

日本国特許庁
JAPAN PATENT OFFICE

310102

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出願年月日
Date of Application:

2001年 6月26日

出願番号
Application Number:

特願2001-192778

出願人
Applicant(s):

沖電気工業株式会社

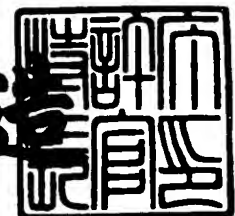


CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年 9月 5日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



出証番号 出証特2001-3081929

【書類名】 特許願

【整理番号】 OG004553

【あて先】 特許庁長官殿

【国際特許分類】 G01L 13/02
G06F 17/21

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 茅原 桂一

【特許出願人】

【識別番号】 000000295

【氏名又は名称】 沖電気工業株式会社

【代理人】

【識別番号】 100089093

【弁理士】

【氏名又は名称】 大西 健治

【手数料の表示】

【予納台帳番号】 004994

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9720320

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 テキスト音声変換装置における高速読み上げ制御方法

【特許請求の範囲】

【請求項1】 入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、

前記パラメータ生成手段は、音韻継続時間を予め経験的に求めた継続時間規則テーブルと、音韻継続時間を統計的手法を用いて予測した継続時間予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記継続時間規則テーブルを用い、閾値を超えていない時には前記継続時間予測テーブルを用いて音韻継続時間の決定を行う音韻継続時間決定手段を有することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項2】 前記閾値は、所定の最大発声速度であることを特徴とする請求項1記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項3】 入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、

前記パラメータ生成手段は、アクセント成分及びフレーズ成分を決定するために必要となるデータを、予め経験的に求めた規則テーブルと、統計的手法を用いて予測した予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記規則テーブルを用い、閾値を超えていない時には前記予測テ

ブルを用いてアクセント成分及びフレーズ成分を決定することによりピッチパターンを決定するピッチパターン決定手段を有することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項4】 前記閾値は、所定の最大発声速度であることを特徴とする請求項3記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項5】 入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、

前記パラメータ生成手段は、前記音声素片を変形させて声質を切り換えるための声質変換係数テーブルを備え、ユーザから指定される発声速度が閾値を超えたときには、声質が変化しないような係数を前記声質変換係数テーブルから選択する声質係数決定手段を有することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項6】 前記閾値は、所定の最大発声速度であることを特徴とする請求項5記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項7】 入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、

前記パラメータ生成手段は、ユーザが指定した抑揚レベルに応じて修正したピッチパターンを出力するピッチパターン修正手段と、ユーザが指定した発声速度に応じて前記修正したピッチパターンを基底ピッチに加算するか否かを選択する切り換

え手段とを有し、前記発声速度が所定の閾値を超えた場合には前記基底ピッチを変更しないように前記切り換え手段を制御することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項 8】 前記閾値は、所定の最大発声速度であることを特徴とする請求項 7 記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項 9】 前記ピッチパターン修正手段は、ユーザが指定した前記発声速度に応じて統計的手法によりフレーズ成分を算出するか或いは当該フレーズ成分を零とする処理を入力文章中に含まれる全フレーズについて行うフレーズ成分算出処理と、ユーザが指定した前記発声速度に応じて統計的手法によりアクセント成分を算出すると共にユーザが指定した前記抑揚レベルに応じて前記算出したアクセント成分を修正するか或いは当該アクセント成分を零とする処理を入力文章中の全ての単語について行う処理とを含むピッチパターン生成処理を行うことを特徴とする請求項 7 記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項 10】 入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、

前記波形生成手段は、文章の切れ目であることを示す合図音を文章間に挿入する合図音生成手段を備え、ユーザから指定される発声速度が閾値を超えた時には前記合図音を文章間に挿入することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項 11】 前記閾値は、所定の最大発声速度であることを特徴とする請求項 10 記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項 12】 入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基

本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、

前記パラメータ生成手段は、ユーザから指定される発声速度が閾値を超えた時には、少なくとも文章の先頭単語の発声速度を通常の発声速度に戻す処理を行う音韻継続時間決定手段を有することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項 1 3】 前記閾値は、所定の最大発声速度であることを特徴とする請求項 1 2 記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項 1 4】 前記音韻継続時間決定手段は、処理中の単語が文章の先頭単語であり、かつユーザから指定される前記発声速度が閾値を超えた時には音韻継続時間を修正しないようにする処理を行い、前記処理中の単語が文章の先頭単語でないか、ユーザから指定された前記発声速度が閾値を超えていない時には、音韻継続時間修正係数をユーザから指定された発声速度に応じて変更する第 1 の処理を行った後、当該単語を構成する母音長を前記音韻継続時間修正係数に従って修正する処理を当該単語の全ての音節について行う第 2 の処理を行い、前記第 1 及び第 2 の処理を当該文章に含まれる全ての単語について行うことを特徴とする請求項 1 2 記載のテキスト音声変換装置における高速読み上げ制御方法。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、日常読み書きしている漢字・仮名混じり文を音声として出力するテキスト音声変換技術に係わり、特に高速読上げ時の韻律制御に関するものである。

【0 0 0 2】

【従来の技術】

テキスト音声変換技術は、我々が日常読み書きしている漢字かな混じり文を入力し、それを音声に変換して出力するもので、出力語彙の制限がないことから録

音・再生型の音声合成に代わる技術として種々の利用分野での応用が期待できる。

従来、この種の音声合成装置としては、図 1 5 に示すような処理形態となっているものが代表的である。

【 0 0 0 3 】

日常読み書きしている漢字仮名混じり文（以下テキストと呼ぶ）を入力すると、テキスト解析部 1 0 1 は、文字情報から音韻・韻律記号列を生成する。ここで、音韻・韻律記号列とは、入力文の読みに加えて、アクセント、イントネーション等の韻律情報を文字列として記述したもの（以下中間言語と呼ぶ）である。単語辞書 1 0 4 は個々の単語の読みやアクセント等が登録された発音辞書で、テキスト解析部 1 0 1 はこの発音辞書を参照しながら、形態素解析ならびに構文解析等の言語処理を施して中間言語を生成する。

【 0 0 0 4 】

テキスト解析部 1 0 1 で生成された中間言語に基づいて、パラメータ生成部 1 0 2 で、音声素片（音の種類）、声質変換係数（声色の種別）、音韻継続時間（音の長さ）、音韻パワー（音の強さ）、基本周波数（声の高さ、以下ピッチと呼ぶ）等の各パタンから成る合成パラメータが決定され、波形生成部 1 0 3 に送られる。

【 0 0 0 5 】

ここで音声素片とは、接続して合成波形を作るための音声の基本単位で、音の種類等に応じて様々なものが用意されている。一般的に、CV、VV、VCV、CVC（C：子音、V：母音）といった音韻連鎖で構成されている場合が多い。

【 0 0 0 6 】

パラメータ生成部 1 0 2 で生成された各種パラメータに基づいて、波形生成部 1 0 3 において音声素片等を蓄積する ROM 等から構成された素片辞書 1 0 5 を参照しながら、合成波形が生成され、スピーカを通して合成音声が出力される。音声合成方法としては、予め音声波形にピッチマーク（基準点）を付けておき、その位置を中心に切り出して、合成時には合成ピッチ周期に合わせて、ピッチマーク位置をずらしながら重ね合わせる方法が知られている。以上がテキスト音声

変換処理の簡単な流れである。

【0007】

次に、パラメータ生成部102における従来の処理を図16を参照して詳細に説明する。

【0008】

パラメータ生成部102に入力される中間言語は、アクセント位置・ポーズ位置などの韻律情報を含んだ音韻文字列であり、これより、ピッチの時間的な変化（以下ピッチパターン）、音声パワー、それぞれの音韻継続時間、素片辞書内に格納されている音声素片アドレス等の波形を生成する上でのパラメータ（以下、総称して合成パラメータと呼ぶ）を決定する。またこの時、ユーザの好みに合わせた発声様式（発声速度、声の高さ、抑揚の大きさ、声の大きさ、発声話者、声質など）を指定するための制御パラメータも入力される場合がある。

【0009】

入力された中間言語に対して、中間言語解析部201で文字列の解析が行われ、中間言語上に記された呼気段落記号・単語区切り記号から単語境界を判定し、アクセント記号からアクセント核のモーラ（音節）位置を得る。呼気段落とは、一息で発声する区間の区切り単位である。アクセント核とは、アクセントが下降する位置のことで、1モーラ目にアクセント核が存在する単語を1型アクセント、nモーラ目にアクセント核が存在する単語をn型アクセントと呼び、総称して起伏型アクセント単語と呼ぶ。逆に、アクセント核の存在しない単語（例えば「新聞」や「パソコン」）を0型アクセントまたは平板型アクセント単語と呼ぶ。これらの韻律に関わる情報は、ピッチパターン決定部202、音韻継続時間決定部203、音韻パワー決定部204、音声素片決定部205、声質係数決定部206に送られる。

【0010】

ピッチパターン決定部202は、中間言語上の韻律情報などからアクセント句あるいはフレーズ単位でのピッチ周波数の時間的な変化パターンの算出を行う。従来では「藤崎モデル」と呼ばれる、臨界制動2次線形系で記述されるピッチ制御機構モデルが用いられてきた。声の高さの情報を与える基本周波数は、次のような過

程で生成されると考えるのがピッチ制御機構モデルである。声帯振動の周波数、すなわち基本周波数は、フレーズの切り替わりごとに発せられるインパルス指令と、アクセントの上げ下げごとに発せられるステップ指令によって制御される。そのとき、生理機構の遅れ特性により、フレーズのインパルス指令は文頭から文末に向かう緩やかな下降曲線（フレーズ成分）となり、アクセントのステップ指令は局所的な起伏の激しい曲線（アクセント成分）となる。これらの二つの成分は、各指令の臨界制動 2 次線形系の応答としてモデル化され、対数基本周波数の時間変化パターンは、これら両成分の和（以降、抑揚成分と呼ぶ）として表現される。

【0011】

図 1.8 はピッチ制御機構モデルを示す。対数基本周波数 $\ln F_0(t)$ (t は時刻) は、次式のように定式化される。

$$\begin{aligned} \ln F_0(t) = & \ln F_{\min} + \sum_{i=1}^I A_{p i} G_{p i}(t - T_{0 i}) \\ & + \sum_{j=1}^J A_{a j} \{ G_{a j}(t - T_{1 j}) - G_{a j}(t - T_{2 j}) \} \quad \dots (1) \end{aligned}$$

ここで、 F_{\min} は最低周波数（以下、基底ピッチと呼ぶ）、 I は文中のフレーズ指令の数、 $A_{p i}$ は文中 i 番目のフレーズ指令の大きさ、 $T_{0 i}$ は文中 i 番目のフレーズ指令の開始時点、 J は文内のアクセント指令の数、 $A_{a j}$ は文内 j 番目のアクセント指令の大きさ、 $T_{1 j}$ 、 $T_{2 j}$ はそれぞれ j 番目のアクセント指令の開始時点と終了時点である。

【0012】

また、 $G_{p i}(t)$ 、 $G_{a j}(t)$ はそれぞれ、フレーズ制御機構のインパルス応答関数、アクセント制御機構のステップ応答関数であり、次式で与えられる。

$$G_{p i}(t) = \alpha_i^2 t \exp(-\alpha_i t) \quad \dots (2)$$

$$G_{a j}(t) = \min [1 - (1 + \beta_j t) \exp(-\beta_j t), \theta] \quad \dots (3)$$

上式は、 $t \geq 0$ の範囲での応答関数であり、 $t < 0$ では $G_{pi}(t) = G_{aj}(t) = 0$ である。式(3)の記号 $\min[x, y]$ は、 x, y のうち小さい方をとることを意味しており、実際の音声でアクセント成分が有限の時間で上限に達することに対応している。ここで、 α_i は i 番目のフレーズ指令に対するフレーズ制御機構の固有角周波数であり、例えば 3.0 など選ばれる。 β_j は j 番目のアクセント指令に対するアクセント制御機構の固有角周波数であり、例えば 20.0 など選ばれる。また、 θ はアクセント成分の上限値であり、例えば 0.9 など選ばれる。

【0013】

なおここで、基本周波数およびピッチ制御パラメータ ($A_{pi}, A_{aj}, T_{0i}, T_{1j}, T_{2j}, \alpha_i, \beta_j, F_{min}$) の値の単位は次のように定義される。すなわち、 $F_0(t)$ および F_{min} の単位は [Hz]、 T_{0i}, T_{1j} および T_{2j} の単位は [sec]、 α_i および β_j の単位は [rad/sec] とする。また A_{pi} および A_{aj} の値は、基本周波数およびピッチ制御パラメータの値の単位を上記のように定めたときの値を用いる。

【0014】

以上で述べた生成過程に基づき、ピッチパターン決定部 202 では、中間言語からピッチ制御パラメータの決定を行う。例えば、フレーズ指令の生起時点 T_{0i} は中間言語上での句読点が存在する位置に設定し、アクセント指令の開始時点 T_{1j} は単語境界記号直後に設定し、アクセント指令の終了時点 T_{2j} はアクセント記号が存在する位置、あるいはアクセント記号がない平板型アクセント単語の場合は、次単語との単語境界記号直前に設定する。フレーズ指令の大きさを表わす A_{pi} とアクセント指令の大きさを表わす A_{aj} は、数量化 I 類などの統計的手法を用いて決定する場合が多い。数量化 I 類については公知であるのでここでは特に説明はしない。

【0015】

図 19 にピッチパターン生成に関する機能ブロック図を示す。中間言語解析部 201 からの解析結果が制御要因設定部 501 に入力される。制御要因設定部 501 では、フレーズ成分、アクセント成分の大きさを予測するために必要な制御要

因の設定を行う。フレーズ成分予測には、例えば、該当するフレーズを構成しているモーラ総数、文内位置、先頭単語のアクセント型といった情報が用いられ、フレーズ成分推定部503に送られる。一方、アクセント成分予測には、例えば、該当するアクセント句のアクセント型、構成しているモーラ総数、品詞、フレーズ内位置といった情報が用いられ、アクセント成分推定部502に送られる。それぞれの成分値予測には、自然発声データを基に数量化I類などの統計的手法を用いて予め学習した予測テーブル506を用いて行われる。

【0016】

予測された結果は、ピッチパターン修正部504に送られ、ユーザから抑揚指定があった場合は、推定された値 A_{pi} 、 A_{aj} に対しての修正を行う。この機能は、文中のある単語を特に強調あるいは抑制したい時に用いることを想定した制御機構である。通常、抑揚指定は3～5段階に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。抑揚指定がない場合は修正は行われない。

【0017】

フレーズ・アクセント両成分値の修正が施された後、基底ピッチ加算部505に送られ、式(1)に従ってピッチパタンの時系列データが生成される。この時、ユーザからの声の高さ指定レベルに従って、基底ピッチテーブル507から指定レベルに応じたデータが基底ピッチとして呼び出され加算される。ユーザから特に指定がない場合は、予め定められたデフォルト値が呼び出され加算される。対数化基底ピッチ $\ln F_{min}$ は合成音声の最低ピッチを表わしており、このパラメータが声の高さの制御に用いられている。通常 $\ln F_{min}$ は、5～10段階に量子化されてテーブルとして保持されておりユーザの好みによって、全体的に声を高くしたい場合は $\ln F_{min}$ を大きくし、逆に声を低くしたい場合は $\ln F_{min}$ を小さくするといった処理を行う。

【0018】

基底ピッチテーブル507は、男声音用と女声音用とに分けられており、ユーザから入力される話者指定によって読み出す基底ピッチを選択する。通常男性音の場合は3.0～4.0の範囲内、女性音の場合は4.0～5.0の範囲内で声

の高さ指定の段階数に応じて量子化されている。以上がピッチパターン生成過程である。

【0019】

次に音韻継続時間制御について述べる。音韻継続時間決定部 203 は、音韻文字列・韻律記号などからそれぞれの音韻の長さ、休止区間長を決定する。休止区間とは、フレーズ間、あるいは文章間でのポーズの長さである（以後ポーズ長と呼ぶ）。音韻長は通常、音節を構成している子音・母音の長さの他、破裂性を有する音韻（p, t, k など）の直前に現れる無音長（閉鎖区間長）を、それぞれ決定する。音韻継続時間長、ポーズ長を総称して継続時間長と呼ぶことにする。音韻継続時間の決定方法は通常、目標となる音韻の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置などにより、数量化 I 類などの統計的手法が用いられる場合が多い。一方、ポーズ長は、前後隣接するフレーズのモーラ総数などにより同じく、数量化 I 類などの統計的手法が用いられる。またこの時、ユーザから発声速度を指定された場合は、それに応じて音韻継続時間の伸縮を行う。通常、発声速度指定は、5～10 段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。音韻継続時間制御に関しては、本発明の主題であるので後述する。

【0020】

音韻パワー決定部 204 は、音韻文字列からそれぞれの音韻の波形振幅値の算出を行う。波形振幅値は、/a, i, u, e, o/ などの音韻の種類・呼気段落内での音節位置などから経験的に決められる。また、音節内においても、立ち上がりの徐々に振幅値が大きくなる区間と、定常状態にある区間と、立ち下がりの徐々に振幅値が小さくなる区間のパワー遷移も同時に決定している。これらパワー制御は通常、テーブル化された係数値を用いることにより実行される。またこの時、ユーザからの声の大きさ指定があった場合は、それに応じて振幅値を増減する。通常、声の大きさ指定は、10 段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。

【0021】

音声素片決定部205は、音韻文字列を表現するために必要な音声素片の、素片辞書105内アドレスの決定を行う。素片辞書105は、例えば男声音と女性音といった具合に複数話者の音声素片が格納されており、ユーザからの話者指定により素片アドレスの決定を行う。素片辞書105に格納されている音声素片データは、CV、VCVなど前後の音韻環境に応じた形で様々な単位で構築されているため、入力テキストの音韻文字列の並びから最適な合成単位を選択する。

【0022】

声質係数決定部206は、ユーザから声質変換指定があった場合に、変換パラメータの決定を行う。声質変換とは、素片辞書105に登録されている素片データに、信号処理等の加工を施すことにより、聴感上、別話者として取り扱えるようにした機能である。一般に、素片データを線形に伸縮する処理を施して実現する場合が多い。伸長処理は、素片データのオーバーサンプリング処理で実現され、太い声となる。逆に縮小処理は、素片データのダウンサンプリング処理で実現され、細い声となる。通常、声質変換指定は、5～10段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられたリサンプリング・レートにより変換を行う。

【0023】

以上の処理により生成されたピッチパターン・音韻パワー・音韻継続時間・音声素片アドレス・伸縮パラメータは合成パラメータ生成部207に送られ、合成パラメータが生成される。合成パラメータは、フレーム（通常8ms程度の長さ）を一つの単位とした波形生成用のパラメータであり、波形生成部103に送られる。

【0024】

図17に波形生成部の機能ブロック図を示す。素片復号部301では、合成パラメータのうち、素片アドレスを参照ポイントとして素片辞書105から素片データをロードし、必要に応じて復号処理を行う。素片辞書105には、音声を合成するための元となる音声素片データが格納されており、何らかの圧縮処理が施されている場合は、復号処理を施す。復号された音素片データは、振幅制御部3

02で振幅係数が乗じられてパワー制御が行われる。素片加工部303では、声質変換のための素片伸縮処理が施される。声質を太くする場合は素片全体を伸長し、声質を細くする場合は素片全体を縮小するといった処理が施される。重畳制御部304では、合成パラメータのうち、ピッチパターンや音韻継続時間といった情報から、素片データの重畳を制御し、合成波形を生成する。波形重畳が完了したデータから逐次DAリングバッファ305に書き込み、出力サンプリング周期でDAコンバータに転送し、スピーカから出力する。

【0025】

次に音韻継続時間制御について詳細に説明する。図20に従来技術による音韻継続時間決定部の機能ブロック図を示す。中間言語解析部201から解析結果が制御要因設定部601に入力される。制御要因設定部601では、例えば、音韻個々の継続時間長あるいは、単語全体での継続時間長などを予測するために必要な制御要因の設定を行う。予測には、例えば、対象となる音韻、前後の音韻の種類、構成しているフレーズのモーラ総数、文内位置といった情報が用いられ、継続時間推定部602に送られる。アクセント成分、フレーズ成分の各成分値予測には、自然発声データを基に数量化I類などの統計的手法を用いて予め学習した継続時間予測テーブル604が用いられる。予測された結果は継続時間修正部603に送られ、ユーザから発声速度指定があった場合は予測値の修正が施される。通常、発声速度指定は、5～10段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。例えば、発声速度レベルが5段階に制御され、レベル0からレベル4まで指定可能だとする。それぞれのレベル n に対応した定数 T_n を次のように定める。すなわち、

$T_0 = 2.0$ 、 $T_1 = 1.5$ 、 $T_2 = 1.0$ 、 $T_3 = 0.75$ 、 $T_4 = 0.5$ とする。

【0026】

先に予測された音韻継続時間のうち、母音長とポーズ長に対して、ユーザから指定されたレベル n に対応した定数 T_n が乗じられる。レベル0の場合は2.0

が乗じられるので生成される波形は長くなり発声速度は遅くなる。レベル4の場合は0.5が乗じられるので生成される波形は短くなり発声速度は速くなる。上記の例では、レベル2が通常発声速度（デフォルト）となっている。

【0027】

発声速度制御が施された合成波形の例を図21に示す。図示したように、音韻継続時間の発声速度制御は通常、母音のみで行う。閉鎖区間長あるいは子音長は、発声速度に依らずほぼ一定と考えられるからである。発声速度を速くした（a）図では母音長だけが0.5倍されており、重畳される音声素片数を減じて実現している。逆に発声速度を遅くした（c）図では母音長だけが1.5倍されており、重畳される音声素片数を繰り返し使うなどして実現している。また、ポーズ長に対しては母音長制御と同様に、指定レベルに応じた定数が乗じられるため、発声速度が遅くなるほどポーズ長も長くなり、発声速度が速くなるほどポーズ長も短くなる。

【0028】

ここで発声速度が速い場合を考える。前述の例ではレベル4に当たる。テキスト音声変換システムの利用特性上、最大発声速度レベルは「早聞き機能」という意味合いが大きい。読上げ対象となるテキストの中でも、ユーザにとって、重要な部分とそうでない部分が存在するため、重要でない部分は発声速度を速くして読み飛ばし、重要な部分は通常発声速度で合成する。このような利用方法が一般的である。最近のテキスト音声変換装置では、早聞き機能用のボタンがあり、このボタンを押下すると発声速度レベルが最大に設定され最高速度で合成され、ボタンを離すと発声速度レベルが以前の設定値に復帰するといったものがある。

【0029】

【発明が解決しようとする課題】

しかしながら上記の従来技術では、以下に述べる問題があった。

（1）早聞き機能を有効にすると、単純に音韻の継続時間長を短くする、言い換えると、生成する波形の長さを短くする処理を施しているため、波形生成部に負荷がかかるといった問題があった。波形生成部では、波形重畳が完了し、生成された波形データから逐次DAリングバッファに書き込むという処理を行っている

ため、生成される波形長が短い場合はその分、波形生成処理に費やすことのできる時間が短くなることになる。波形データ長が半分になると、処理時間も半分で終了させなければならない。例えば、音韻継続時間長が半分になったからといって、必ずしも演算量が半分になるわけではないため、DAコンバータへの転送処理に、波形生成処理が追いつかない場合は、合成音が途中で止まる「音切れ」現象が発生する場合がある。

【0030】

(2) 早聞き機能を有効にすると、単純に音韻の継続時間長を短くする処理が施されるため、ピッチパターンも基本的に線形に縮小される。つまり抑揚も時間的に速い周期で変動することになり、これは、不自然なイントネーションで非常に聞き取りにくい合成音となっていた。早聞き機能は、読上げ対象となるテキストを完全にスキップするのではなく、聞き流すという用途で用いられるため、抑揚の激しい合成音は不向きであった。従来技術において早聞き機能有効時の合成音声は、抑揚変化が激しすぎるため聞き取りにくく理解しづらいものとなっていた。

【0031】

(3) 早聞き機能を有効にすると、音韻継続時間と共に、文章間のポーズも同一比率で縮小される。そのため、文章と文章の境界がほとんどなくなり、切れ目が分かり難くなっていた。1文の合成音声を出力した直後に、さらに次の1文の合成音声が出力されるため、従来技術において早聞き機能有効時の合成音声は、テキスト内容を理解しつつ読み飛ばす用途においては不向きであった。

【0032】

(4) 早聞き機能を有効にすると、テキスト全体に渡って、発声速度が速くなるため、早聞き解除のタイミングを取ることが難しかった。通常の早聞き機能使用方法は、ある文章の中から所望の部分までを読み飛ばし、以降を通常速度で合成するというものである。従来技術によると、ユーザが欲した部分の読上げが行われ、早聞き機能解除をした時点では、所望の部分を大きく通り越してしまうといった問題があった。この場合、早聞き機能を解除した後一旦、読上げ対象区間を前にさかのぼって設定した後に通常発声速度で合成開始するといった面倒な操作をしなければいけなかった。またユーザは、必要な部分と必要でない部分とを聞

き分けながら、早聞き機能の有効化・無効化の動作を行わなければならない、非常に労力を必要としていた。

【 0 0 3 3 】

本発明は、（A）発声速度を速くした時に高負荷になって音切れが発生するという問題点と、（B）発声速度を速くした時にピッチ変動周期も速くなり、不自然なイントネーションになってしまうという問題点と、（C）発声速度を速くした時に、文と文との間のポーズが短くなって理解しがたい合成音になってしまうという問題点と、（D）早聞き機能解除タイミングが取りづらいという問題点を解決したテキスト音声変換における高速読み上げ制御方法を提供することを目的とする。

【 0 0 3 4 】

【課題を解決するための手段】

この発明は、上記課題（A）を解決するために、ユーザの指定する発声速度が最高速に設定された場合、すなわち早聞き機能が有効となった場合に、パラメータ生成手段における音韻継続時間決定手段において、統計的手法を用いて予測した継続時間予測テーブルに替えて、予め経験的に求めた継続時間規則テーブルを用いて音韻継続時間を決定し、また、ピッチパターン決定手段において、統計的手法により算出した予測テーブルを用いる代わりに、予め経験的に求めた規則テーブルを使用してピッチパターンを決定し、更に、声質決定手段においては声質が変化しないような声質変換係数を選択する。

【 0 0 3 5 】

また、この発明は、上記課題（B）を解決するために、ユーザの指定する発声速度が最高速に設定された場合に、アクセント成分及びフレーズ成分の計算を行わないようにすると共に基底ピッチを変更しないようにしている。

【 0 0 3 6 】

また、この発明は、上記課題（C）を解決するために、ユーザの指定する発声速度が最高速に設定された場合に、文章間に文章の区切りを示す合図音を挿入するようにしている。

【 0 0 3 7 】

更に、この発明は、上記課題（D）を解決するために、ユーザの指定する発声速度が最高速に設定された場合でも、少なくとも文章の先頭単語については、通常の発声速度で再生するようにしている。

【0038】

【発明の実施の形態】

第1の実施の形態

〔構成〕

以下、第1の実施の形態における構成を図面を参照しながら詳細に説明する。従来技術と異なる点は、発声速度が最高速に設定された場合、すなわち、早聞き機能が有効となった場合に内部演算処理の一部を簡略化、省略を行うことによって負荷軽減させた点である。

【0039】

図1は、第1の実施の形態におけるパラメータ生成部102の機能ブロック図である。パラメータ生成部102への入力は従来と同じく、テキスト解析部101から出力される中間言語および、ユーザが個別に指定する韻律制御パラメータである。中間言語解析部801には一文毎の中間言語が入力され、以降の韻律生成処理で必要となる音韻系列・フレーズ情報・アクセント情報などといった中間言語解析結果が、それぞれピッチパターン決定部802、音韻継続時間決定部803、音韻パワー決定部804、音声素片決定部805、声質係数決定部806に出力される。

【0040】

ピッチパターン決定部802には、前述の中間言語解析結果に加えてユーザからの抑揚指定・声の高さ指定・発声速度指定・話者指定の各パラメータが入力され、ピッチパターンが合成パラメータ生成部807に出力される。ピッチパターンとは基本周波数の時間的遷移のことである。

【0041】

音韻継続時間決定部803には、前述の中間言語解析結果に加えてユーザからの発声速度指定のパラメータが入力され、それぞれの音韻の音韻継続時間・ポーズ長といったデータが合成パラメータ生成部807に出力される。

【0042】

音韻パワー決定部804には、前述の中間言語解析結果に加えてユーザからの声の大きさ指定パラメータが入力され、それぞれの音韻の音韻振幅係数が合成パラメータ生成部807に出力される。

【0043】

音声素片決定部805には、前述の中間言語解析結果に加えてユーザからの話者指定パラメータが入力され、波形重畳するための必要な音声素片アドレスが合成パラメータ生成部807に出力される。

【0044】

声質係数決定部806には、前述の中間言語解析結果に加えてユーザからの声質指定・発声速度指定の各パラメータが入力され、声質変換パラメータが合成パラメータ生成部807に出力される。

【0045】

合成パラメータ生成部807は、入力された各韻律パラメータ（前述したピッチパターン、音韻継続時間、ポーズ長、音韻振幅係数、音声素片アドレス、声質変換係数）から、フレーム（通常8ms程度の長さ）を一つの単位とした波形生成用のパラメータを生成し、波形生成部103に出力する。

【0046】

パラメータ生成部102において、従来技術と比較して異なる点は、発声速度指定パラメータが音韻継続時間決定部803のほかに、ピッチパターン決定部802、声質係数決定部806のそれぞれに入力されている点と、ピッチパターン決定部802、音韻継続時間決定部803、声質係数決定部806のそれぞれの内部処理である。テキスト解析部101および波形生成部103においては、従来と同様であるため、その構成に関する説明は省略する。

【0047】

ピッチパターン決定部802の構成について図2を用いて説明する。第1の実施の形態においては、アクセント成分およびフレーズ成分の決定に、数量化I類等の統計的手法を用いる場合と規則による場合との2通りの構成を有する。規則による制御の場合は、予め経験的に求められた規則テーブル910を用い、統計的

手法による制御の場合は、自然発声データを基に数量化Ⅰ類などの統計的手法を用いて予め学習した予測テーブル 9 0 9 を用いる。予測テーブル 9 0 9 のデータ出力はスイッチ 9 0 7 の a 端子に接続され、規則テーブル 9 1 0 のデータ出力はスイッチ 9 0 7 の b 端子に接続される。いずれの端子が選択されるかは、セレクト 9 0 6 の出力によって決定される。

【 0 0 4 8 】

セレクト 9 0 6 には、ユーザから指定される発声速度レベルが入力され、スイッチ 9 0 7 を制御するための信号がスイッチ 9 0 7 に接続される。発声速度が最高レベルの場合はスイッチ 9 0 7 を b 端子側に接続し、それ以外の場合はスイッチ 9 0 7 を a 端子側に接続する。スイッチ 9 0 7 の出力は、アクセント成分決定部 9 0 2 とフレーズ成分決定部 9 0 3 に接続される。

【 0 0 4 9 】

中間言語解析部 8 0 1 からの出力は制御要因設定部 9 0 1 に入力され、アクセント・フレーズ両成分の決定のための要因パラメータの解析が行われ、その出力がアクセント成分決定部 9 0 2 とフレーズ成分決定部 9 0 3 に接続される。

【 0 0 5 0 】

アクセント成分決定部 9 0 2 とフレーズ成分決定部 9 0 3 には、スイッチ 9 0 7 からの出力が接続されており、予測テーブル 9 0 9 もしくは規則テーブル 9 1 0 を用いてそれぞれの成分値を決定しピッチパターン修正部 9 0 4 に出力する。

【 0 0 5 1 】

ピッチパターン修正部 9 0 4 には、ユーザから指定される抑揚指定レベルが入力され、該レベルに応じて予め定められた定数が乗じられ、その結果が基底ピッチ加算部 9 0 5 に接続される。

【 0 0 5 2 】

基底ピッチ加算部 9 0 5 にはさらに、ユーザから指定される声の高さレベル・話者指定および、基底ピッチテーブル 9 0 8 が接続されている。基底ピッチテーブル 9 0 8 には、ユーザ指定された声の高さレベルと性別とに応じて予め定められた定数値が格納されており、ピッチパターン修正部 9 0 4 からの入力に加算してピッチパターン時系列データとして合成パラメータ生成部 8 0 7 に出力する。

【0053】

音韻継続時間決定部 803 の構成について図 3 を用いて説明する。第 1 の実施の形態においては、音韻継続時間の決定に、数量化 I 類等の統計的手法を用いる場合と規則による場合との 2通りの構成を有する。規則による制御の場合は、予め経験的に求められた継続時間規則テーブル 1007 を用い、統計的手法による制御の場合は、自然発声データを基に数量化 I 類などの統計的手法を用いて予め学習した継続時間予測テーブル 1006 を用いる。継続時間予測テーブル 1006 のデータ出力はスイッチ 1005 の a 端子に接続され、継続時間規則テーブル 1007 のデータ出力はスイッチ 1005 の b 端子に接続される。いずれの端子が選択されるかは、セレクト 1004 の出力によって決定される。

【0054】

セレクト 1004 には、ユーザから指定される発声速度レベルが入力され、スイッチ 1005 を制御するための信号がスイッチ 1005 に接続される。発声速度が最高レベルの場合はスイッチ 1005 を b 端子側に接続し、それ以外の場合はスイッチ 1005 を a 端子側に接続する。スイッチ 1005 の出力は、継続時間決定部 1002 に接続される。

【0055】

中間言語解析部 801 からの出力は制御要因設定部 1001 に入力され、音韻継続時間決定のための要因パラメータの解析が行われ、その出力が継続時間決定部 1002 に接続される。

【0056】

継続時間決定部 1002 には、スイッチ 1005 からの出力が接続されており、継続時間予測テーブル 1006 もしくは継続時間規則テーブル 1007 を用いて音韻継続時間長を決定し継続時間修正部 1003 に出力する。継続時間修正部 1003 には、ユーザから指定される発声速度レベルが入力され、該レベルに応じて予め定められた定数が乗じられて修正が施され、その結果が合成パラメータ生成部 807 に出力される。

【0057】

声質係数決定部 806 の構成について図 4 を用いて説明する。この例では声質

変換指定レベルは 5 段階となっている。ユーザから指定される発声速度レベルおよび声質指定レベルがセクタ 1 1 0 2 に入力され、スイッチ 1 1 0 3 を制御するための信号がスイッチ 1 1 0 3 に接続される。この時のスイッチ制御信号は、発声速度が最高レベルの場合は無条件で c 端子有効にし、それ以外の場合は、声質指定レベルに応じた端子が有効となる。すなわち、声質レベルが 0 の時は a 端子、レベル 1 の時は b 端子、以下同様にレベル 4 の時 e 端子がそれぞれ有効となる。スイッチ 1 1 0 3 の a ～ e の各端子は、声質変換係数テーブル 1 1 0 4 に接続され、それぞれに対応した声質変換係数データが呼び出され、スイッチ 1 1 0 3 の出力として声質係数選択部 1 1 0 1 に接続される。声質係数選択部 1 1 0 1 は入力された声質変換係数を合成パラメータ生成部 8 0 7 に出力する。

【 0 0 5 8 】

〔動作〕

以上のように構成された第 1 の実施の形態における動作について詳細に説明する。従来技術と異なる点は、パラメータ生成に関わる処理であるので、それ以外の処理については説明を省略する。

【 0 0 5 9 】

テキスト解析部 1 0 1 で生成された中間言語は、パラメータ生成部 1 0 2 内部の中間言語解析部 8 0 1 に送られる。中間言語解析部 8 0 1 では、中間言語上に記述されているフレーズ区切り記号、単語区切り記号、アクセント核を示すアクセント記号、そして音韻記号列から、韻律生成に必要なデータを抽出して、ピッチパターン決定部 8 0 2、音韻継続時間決定部 8 0 3、音韻パワー決定部 8 0 4、音声素片決定部 8 0 5、声質係数決定部 8 0 6 のそれぞれの機能ブロックへ送る。

【 0 0 6 0 】

ピッチパターン決定部 8 0 2 では、声の高さの遷移であるイントネーションが生成され、音韻継続時間決定部 8 0 3 では、音韻個々の継続時間のほか、フレーズとフレーズの切れ目あるいは、文と文との切れ目に挿入するポーズ長を決定する。また、音韻パワー決定部 8 0 4 では、音声波形の振幅値の遷移である音韻パワーが生成され、音声素片決定部 8 0 5 では合成波形を生成するために必要となる音

声素片の、素片辞書 1 0 5 におけるアドレスを決定する。声質係数決定部 8 0 6 では、素片データを信号処理で加工するためのパラメータの決定が行われる。ユーザから指定される韻律制御指定のうち、抑揚指定および声の高さ指定はピッチパターン決定部 8 0 2 に、発声速度指定はピッチパターン決定部 8 0 2 と音韻継続時間決定部 8 0 3 と声質係数決定部 8 0 6 に、声の大きさ指定は音韻パワー決定部 8 0 4 に、話者指定はピッチパターン決定部 8 0 2 と音声素片決定部 8 0 5 に、声質指定は声質係数決定部 8 0 6 にそれぞれ送られている。

【 0 0 6 1 】

以下に、それぞれの機能ブロックごとに動作の説明を行う。

まず、図 2 を用いて、ピッチパターン決定部 8 0 2 の動作を詳細に説明する。中間言語解析部 2 0 1 から解析結果が制御要因設定部 9 0 1 に入力される。制御要因設定部 9 0 1 では、フレーズ成分、アクセント成分の大きさを決定するために必要な制御要因の設定を行う。フレーズ成分の大きさの決定に必要なデータとは、例えば、該当するフレーズを構成しているモーラ総数、文内での相対位置、先頭単語のアクセント型といった情報である。一方、アクセント成分の大きさの決定に必要なデータとは、例えば、該当するアクセント句のアクセント型、構成しているモーラ総数、品詞、フレーズ内での相対位置といった情報である。これらの成分値を決定するために予測テーブル 9 0 9 あるいは、規則テーブル 9 1 0 が使用される。前者は、自然発声データを基に数量化 I 類などの統計的手法を用いて予め学習したテーブルであり、後者は、予備実験等の実施により経験的に導き出された成分値が格納されたテーブルである。数量化 I 類に関しては公知であるのでここでは説明を省略する。どちらが選択されるかはスイッチ 9 0 7 により制御され、スイッチ 9 0 7 が a 端子に接続された場合は予測テーブル 9 0 9 が、b 端子に接続された場合は規則テーブル 9 1 0 が選択されることになる。

【 0 0 6 2 】

ピッチパターン決定部 8 0 2 には、ユーザから指定される発声速度レベルが入力されており、これによりセレクタ 9 0 6 を介してスイッチ 9 0 7 が駆動されている。セレクタ 9 0 6 は、入力された発声速度レベルが最高速度であった時、スイッチ 9 0 7 を b 端子側に接続するような制御信号を送信する。逆に、入力された

発声速度レベルが最高速度ではない時、スイッチ 9 0 7 を a 端子側に接続するような制御信号を送信する。例えば、発声速度が 5 段階、レベル 0 からレベル 4 まですべて設定でき、数値が大きくなる程発声速度が速くなる仕様の場合、セレクト 9 0 6 は、入力された発声速度レベルが 4 の時だけスイッチ 9 0 7 を b 端子に接続するような制御信号を送信し、それ以外の場合は a 端子に接続するような制御信号を送信する。すなわち、発声速度が最高速度の場合は規則テーブル 9 1 0 が選択され、そうでない場合は予測テーブル 9 0 9 が選択されることになる。

【 0 0 6 3 】

アクセント成分決定部 9 0 2 とフレーズ成分決定部 9 0 3 は、選択されたテーブルを用いてそれぞれの成分値の算出を行う。予測テーブル 9 0 9 が選択された場合は、統計的手法を用いてアクセント・フレーズ両成分の大きさを決定する。規則テーブル 9 1 0 が選択された場合は、あらかじめ決められた規則に従ってアクセント・フレーズ両成分の大きさを決定する。例えばフレーズ成分の大きさの規則化の例としては、文内の位置で決定し、文先頭フレーズは一律に 0. 3、文終端フレーズは一律に 0. 1、それ以外の文中フレーズは 0. 2 などが考えられる。アクセント成分の大きさに関しても、アクセント型が 1 型の時とそれ以外の場合、フレーズ内での単語位置が先頭の場合とそうでない場合といった具合に場合分けして、それぞれの条件に対して成分値を割り当てておく。このような構成にすることで、フレーズ・アクセント両成分値の決定はテーブル参照を行うだけで行える。本発明におけるピッチパターン決定部の主題は、統計的手法を用いてフレーズ・アクセント成分の大きさを決定する場合と比較して、演算量が少なく済み、処理時間の短縮が図れるモードを有する構成にすることである。したがって、規則化手順は上記に限られるものではない。

【 0 0 6 4 】

以上のような処理が施され決定したアクセント成分、フレーズ成分は、ピッチパターン修正部 9 0 4 で抑揚制御が行われ、基底ピッチ加算部 9 0 5 で声の高さ制御が施される。

【 0 0 6 5 】

ピッチパターン修正部 9 0 4 はユーザから指定される抑揚制御レベルに応じた係

数を乗ずる操作が行われる。ユーザからの抑揚制御指定は例えば、3段階で与えられ、レベル1が抑揚を1.5倍に、レベル2が抑揚を1.0倍に、レベル3が抑揚を0.5倍にといった具合に定められている。

【0066】

基底ピッチ加算部905では、抑揚修正されたアクセント成分、フレーズ成分に対して、ユーザから指定される声の高さレベルあるいは、話者指定（性別）に応じた定数を加算する操作が行われ、ピッチパターン時系列データとして合成パラメータ生成部807に送られる。例えば、声の高さレベルが5段階、レベル0からレベル4まで設定できるシステムの場合、基底ピッチテーブル908に格納されているデータは男声音の場合、3.0、3.2、3.4、3.6、3.8といった数値、女性音の場合は、4.0、4.2、4.4、4.6、4.8といった数値が良く用いられる。

【0067】

次に音韻継続時間制御について図3を用いてその動作について詳細に説明する。中間言語解析部201から解析結果が制御要因設定部1001に入力される。制御要因設定部1001では、音韻継続時間（子音長・母音長・閉鎖区間長）、ポーズ長を決定するために必要な制御要因の設定を行う。音韻継続時間の決定に必要なデータとは、例えば、目標となる音韻の種別、対象音節の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置といった情報である。一方、ポーズ長決定に必要なデータとは、前後隣接するフレーズのモーラ総数といった情報である。これらの継続時間長を決定するために継続時間予測テーブル1006あるいは、継続時間規則テーブル1007が使用される。前者は、自然発声データを基に数量化I類などの統計的手法を用いて予め学習したテーブルであり、後者は、予備実験等の実施により経験的に導き出された成分値が格納されたテーブルである。どちらが選択されるかはスイッチ1005により制御され、スイッチ1005がa端子に接続された場合は継続時間予測テーブル1006が、b端子に接続された場合は継続時間規則テーブル1007が選択されることになる。

【0068】

音韻継続時間決定部803には、ユーザから指定される発声速度レベルが入力

されており、これによりセクタ1004を介してスイッチ1005が駆動されている。セクタ1004は、入力された発声速度レベルが最高速度であった時、スイッチ1005をb端子側に接続するような制御信号を送信する。逆に、入力された発声速度レベルが最高速度ではない時は、スイッチ1005をa端子側に接続するような制御信号を送信する。例えば、発声速度が5段階、レベル0からレベル4まで設定でき、数値が大きくなる程発声速度が速くなる仕様の場合、セクタ1004は、入力された発声速度レベルが4の時だけスイッチ1005をb端子に接続するような制御信号を送信し、それ以外の時はa端子に接続するような制御信号を送信する。すなわち、発声速度が最高速度の場合は継続時間規則テーブル1007が選択され、そうでない場合は継続時間予測テーブル1006が選択されることになる。

【0069】

継続時間決定部1002は、選択されたテーブルを用いて音韻継続時間、ポーズ長の算出を行う。継続時間予測テーブル1006が選択された場合は、統計的手法を用いて決定する。継続時間規則テーブル1007が選択された場合は、あらかじめ決められた規則に従って決定する。例えば音韻継続時間の規則化の例としては、その音韻の種類、文内の位置などに応じて基本長を割り当てておく。大量の自然発声データから音韻毎に平均を算出し、これを基本長としてもよい。ポーズ長に関しては、一律に300msを割り当てるか、あるいは、テーブル参照を行うだけで決定できるような構成が望ましい。本実施の形態における音韻継続時間決定部の主題は、統計的手法を用いて継続時間を決定する場合と比較して、演算量が少なく済み、処理時間の短縮が図れるモードを有する構成にすることである。したがって、規則化手順は上記に限られるものではない。

【0070】

以上のような処理が施され決定した継続時間は、継続時間修正部1003に送られる。継続時間修正部1003には、ユーザから指定される発声速度レベルも同時に入力されており、このレベルに応じて音韻継続時間の伸縮を行う。通常、発声速度指定は、5～10段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を母音の継続時間長あるいは、ポーズ長に対して乗ず

ることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。

【0071】

次に声質係数決定について図4を用いてその動作について詳細に説明する。声質係数決定部806には、ユーザから指定される声質変換レベルと、発声速度レベルが入力される。これらの韻律制御パラメータは、セレクトタ1102を介してスイッチ1103を制御するために用いられる。セレクトタ1102はまず、発声速度レベルの判定を行う。発声速度レベルが最高速度の場合は、スイッチ1103をc端子に接続し、最高速度以外の場合は、声質変換レベルの判定を行う。この時は、声質変換レベルに応じた端子に接続するようにスイッチ1103を制御する。声質指定レベルが0の時はa端子、レベル1の時はb端子、以下同様にレベル4の時はe端子に接続する。スイッチ1103のa～eの各端子は、声質変換係数テーブル1104に接続され、それぞれに対応した声質変換係数データが呼び出される機能になっている。

【0072】

声質変換係数テーブル1104には、音声素片の伸縮係数が格納されており、例えば声質変換レベルnに対応する伸縮係数を K_n を次のように定める。すなわち、

$K_0 = 2.0$ 、 $K_1 = 1.5$ 、 $K_2 = 1.0$ 、 $K_3 = 0.8$ 、 $K_4 = 0.5$ のように設定する。これらの数値は、元となる音声素片の長さを K_n 倍に伸縮した後に波形重畳して合成音声を生成するという意味である。レベル2の時は、係数値が1.0なので声質変換のための処理は一切行われなことになる。スイッチ1103のa端子に接続されている場合は、係数 K_0 が選択されて声質係数選択部1101に送られる。スイッチ1103のb端子に接続されている場合は、係数 K_1 が選択されて声質係数選択部1101に送られるといった具合である。

【0073】

ここで、図5を参照しながら素片の線形伸縮の方法の一例について述べる。声質変換レベルnにおける音声素片のデータの第mサンプル目を X_{nm} とする。このように定義すると、声質変換後のデータ系列は、変換前のデータ系列 X_{2n} を

用いて以下のようにして算出することができる。即ち、

レベル0では、

$$\begin{aligned} X_{00} &= X_{20} \\ X_{01} &= X_{20} \times 1/2 + X_{21} \times 1/2 \\ X_{02} &= X_{21} \end{aligned}$$

レベル1では、

$$\begin{aligned} X_{10} &= X_{20} \\ X_{11} &= X_{20} \times 1/3 + X_{21} \times 2/3 \\ X_{12} &= X_{21} \times 2/3 + X_{22} \times 1/3 \\ X_{13} &= X_{22} \end{aligned}$$

レベル3では、

$$\begin{aligned} X_{30} &= X_{20} \\ X_{31} &= X_{21} \times 3/4 + X_{22} \times 1/4 \\ X_{32} &= X_{22} \times 1/2 + X_{23} \times 1/2 \\ X_{33} &= X_{23} \times 1/4 + X_{24} \times 3/4 \\ X_{34} &= X_{25} \end{aligned}$$

レベル4では、

$$\begin{aligned} X_{40} &= X_{20} \\ X_{41} &= X_{22} \end{aligned}$$

のようになる。上記は、声質変換のための一例であって、これに限られるものではない。本実施の形態における声質係数決定部の主題は、発声速度レベルが最高速の時に声質変換指定を無効とする機能を有することにより、処理時間の短縮を図ることである。

【0074】

以上詳細に説明したように、第1の実施の形態によれば、発声速度が既定値最大に設定された場合に、テキスト音声変換処理の中で演算負荷が大きい機能ブロックを簡略化あるいは、無効にする処理を施しているため、高負荷による音切れが発生する機会を減少させ、聞き易い合成音声を生成することが可能となる。

【0075】

この場合、発声速度が最高レベル以外に設定された時の合成音と比較して、ピッチや継続時間などの韻律性能の若干の違い、声質変換機能が有効とならない、といったことが起きるが、最高速度での合成音出力は通常、読み飛ばしという意味合いで利用される場合がほとんどある。したがって、音声出力されるテキストの内容を把握・理解できれば良い、という程度の使用方法なので声質変換機能の有無、あるいは韻律性能低下といった点は音切れ現象と比較すると許容できるものと考えられる。

【0076】

第2の実施の形態

〔構成〕

第2の実施の形態における構成を図面を参照しながら詳細に説明する。本実施の形態が従来技術と異なる点は、発声速度が最高速に設定された場合、すなわち、早聞き機能が有効となった時にピッチパターン生成処理を変更する点である。したがって、従来と異なるパラメータ生成部、ピッチパターン決定部についてのみ説明する。

【0077】

図6は第2の実施の形態におけるパラメータ生成部の機能ブロック図を示しており、このブロック図を用いて説明する。パラメータ生成部102への入力とは従来と同じく、テキスト解析部101から出力される中間言語および、ユーザが個別に指定する韻律制御パラメータである。中間言語解析部1301には一文毎の中間言語が入力され、以降の韻律生成処理で必要となる音韻系列・フレーズ情報・アクセント情報などといった中間言語解析結果が、それぞれピッチパターン決定部1302、音韻継続時間決定部1303、音韻パワー決定部1304、音声素片決定部1305、声質係数決定部1306に出力される。

【0078】

ピッチパターン決定部1302には、前述の中間言語解析結果に加えてユーザからの抑揚指定・声の高さ指定・発声速度指定・話者指定の各パラメータが入力され、ピッチパターンが合成パラメータ生成部1307に出力される。

【0079】

音韻継続時間決定部 1303 には、前述の中間言語解析結果に加えてユーザからの発声速度指定のパラメータが入力され、それぞれの音韻継続時間・ポーズ長といったデータが合成パラメータ生成部 1307 に出力される。

【0080】

音韻パワー決定部 1304 には、前述の中間言語解析結果に加えてユーザからの声の大きさ指定パラメータが入力され、それぞれの音韻振幅係数が合成パラメータ生成部 1307 に出力される。

【0081】

音声素片決定部 1305 には、前述の中間言語解析結果に加えてユーザからの話者指定パラメータが入力され、波形重畳するための必要な音声素片アドレスが合成パラメータ生成部 1307 に出力される。

【0082】

声質係数決定部 1306 には、前述の中間言語解析結果に加えてユーザからの声質指定・発声速度指定の各パラメータが入力され、声質変換パラメータが合成パラメータ生成部 1307 に出力される。

【0083】

合成パラメータ生成部 1307 は、入力された各韻律パラメータ（前述したピッチパターン、音韻継続時間、ポーズ長、音韻振幅係数、音声素片アドレス、声質変換係数）を、フレーム（通常 8 ms 程度の長さ）を一つの単位とした波形生成用のパラメータに変換し、波形生成部 103 に出力する。

【0084】

パラメータ生成部 102 において、従来技術と比較して異なる点は、発声速度指定パラメータが音韻継続時間決定部 1303 のほかに、ピッチパターン決定部 1302 に入力されている点と、ピッチパターン決定部 1302 の内部処理である。テキスト解析部 101 および波形生成部 103 においては、従来と同様であるため、その構成に関する説明は省略する。また、パラメータ生成部 102 の内部機能ブロックにおいても、ピッチパターン決定部 1302 以外は従来と同様であるため、その構成に関する説明は省略する。

【0085】

ピッチパターン決定部 1 3 0 2 の構成について図 7 を用いて説明する。中間言語解析部 1 3 0 1 からの出力は制御要因設定部 1 4 0 1 に入力され、アクセント・フレーズ両成分の決定のための要因パラメータの解析が行われ、その出力がアクセント成分決定部 1 4 0 2 とフレーズ成分決定部 1 4 0 3 に接続される。

【 0 0 8 6 】

アクセント成分決定部 1 4 0 2 とフレーズ成分決定部 1 4 0 3 には、予測テーブル 1 4 0 8 が接続され、数量化 I 類等の統計的手法を用いてそれぞれの成分の大きさを予測する。予測されたアクセント成分値、フレーズ成分値はピッチパターン修正部 1 4 0 4 に接続される。

【 0 0 8 7 】

ピッチパターン修正部 1 4 0 4 にはユーザから指定される抑揚指定レベルが入力され、該レベルに応じて予め定められた定数が前述のアクセント成分、フレーズ成分に乘じられ、その結果がスイッチ 1 4 0 5 の a 端子に接続される。スイッチ 1 4 0 5 にはさらに b 端子が存在し、セレクトア 1 4 0 6 から出力される制御信号により、端子 a、端子 b のいずれかに接続されるように構成されている。

【 0 0 8 8 】

セレクトア 1 4 0 6 には、ユーザから指定される発声速度レベルが入力され、発声速度が最高レベルの場合はスイッチ 1 4 0 5 を b 端子に接続し、それ以外の場合はスイッチ 1 4 0 5 を a 端子に接続する制御信号を出力する。スイッチ 1 4 0 5 の b 端子は常にグランドに接続されており、スイッチ 1 4 0 5 は、a 端子が有効の時はピッチパターン修正部 1 4 0 4 からの出力を、b 端子が有効の時は 0 を基底ピッチ加算部 1 4 0 7 に出力する機能を有している。

【 0 0 8 9 】

基底ピッチ加算部 1 4 0 7 にはさらに、ユーザから指定される声の高さレベル・話者指定および、基底ピッチテーブル 1 4 0 9 が接続されている。基底ピッチテーブル 1 4 0 9 には、ユーザ指定された声の高さレベルと話者の性別に応じて予め定められた定数値が格納されており、スイッチ 1 4 0 5 からの入力に加算してピッチパターン時系列データとして合成パラメータ生成部 1 3 0 7 に出力する。

【 0 0 9 0 】

【動作】

以上のように構成された本発明の第2の実施の形態における動作について詳細に説明する。

【0091】

まず、テキスト解析部101で生成された中間言語は、パラメータ生成部102内部の中間言語解析部1301に送られる。中間言語解析部1301では、中間言語上に記述されているフレーズ区切り記号、単語区切り記号、アクセント核を示すアクセント記号、そして音韻記号列から、韻律生成に必要なデータを抽出して、ピッチパターン決定部1302、音韻継続時間決定部1303、音韻パワー決定部1304、音声素片決定部1305、声質係数決定部1306のそれぞれの機能ブロックへ送る。

【0092】

ピッチパターン決定部1302では、声の高さの遷移であるイントネーションが生成され、音韻継続時間決定部1303では、音韻個々の継続時間のほか、フレーズとフレーズの切れ目あるいは、文と文との切れ目に挿入するポーズ長を決定する。また、音韻パワー決定部1304では、音声波形の振幅値の遷移である音韻パワーが生成され、音声素片決定部1305では合成波形を生成するために必要となる音声素片の、素片辞書105におけるアドレスを決定する。声質係数決定部1306では、素片データを信号処理で加工するためのパラメータの決定が行われる。

【0093】

ユーザから指定される種々の韻律制御指定のうち、抑揚指定および声の高さ指定はピッチパターン決定部1302に、発声速度指定はピッチパターン決定部1302と音韻継続時間決定部1303に、声の大きさ指定は音韻パワー決定部1304に、話者指定はピッチパターン決定部1302と音声素片決定部1305に、声質指定は声質係数決定部1306にそれぞれ送られている。

【0094】

以下に図7を用いてピッチパターン決定部1302の動作に関して説明する。従来技術と異なる点は、ピッチパターン生成に関わる処理であるので、それ以外の処

理については省略する。

【0095】

中間言語解析部201から解析結果が制御要因設定部1401に入力される。制御要因設定部1401では、フレーズ成分、アクセント成分の大きさを予測するために必要な制御要因の設定を行う。フレーズ成分の大きさの予測に必要なデータとは、例えば、該当するフレーズを構成しているモーラ総数、文内での相対位置、先頭単語のアクセント型といった情報である。一方、アクセント成分の大きさの予測に必要なデータとは、例えば、該当するアクセント句のアクセント型、構成しているモーラ総数、品詞、フレーズ内での相対位置といった情報である。これらの成分値を決定するために予測テーブル1408が使用される。予測テーブル1408は、自然発声データを基に数量化I類などの統計的手法を用いて予め学習したテーブルである。数量化I類に関しては公知であるのでここでは説明を省略する。

【0096】

制御要因設定部1401で解析された予測制御要因は、アクセント成分決定部1402とフレーズ成分決定部1403に送られ、それぞれにおいてアクセント成分の大きさ、フレーズ成分の大きさが予測テーブル1408を用いて予測される。第1の実施の形態でも示したように、予測モデルを使わずに規則でそれぞれの成分値を決定しても構わない。算出されたアクセント成分、フレーズ成分は、ピッチパターン修正部1404に送られ、ユーザから指定される抑揚指定レベルに応じた係数を乗ずる操作が行われる。

【0097】

ユーザからの抑揚制御指定は例えば、3段階で与えられ、レベル1が抑揚を1.5倍に、レベル2が抑揚を1.0倍に、レベル3が抑揚を0.5倍にといった具合に定められている。

【0098】

修正されたアクセント、フレーズ両成分はスイッチ1405のa端子に送られる。スイッチ1405は、a、b、2つの端子を有しており、セレクト1406からの制御信号によりどちらかの端子に接続するような機能になっている。一方

のb端子は常に0が入力されるようになっている。

【0099】

セレクト1406にはユーザからの発声速度レベルが入力されており、これにより出力制御が行われている。セレクト1406は、入力された発声速度レベルが最高速度であった時、スイッチ1405をb端子側に接続するような制御信号を送信する。逆に、入力された発声速度レベルが最高速度ではない時、スイッチ1405をa端子側に接続するような制御信号を送信する。例えば、発声速度が5段階、レベル0からレベル4まで設定でき、数値が大きくなる程発声速度が速くなる仕様の場合、セレクト1406は、入力された発声速度レベルが4の時だけスイッチ1405をb端子に接続するような制御信号を送信し、それ以外の時はa端子に接続するような制御信号を送信する。すなわち、発声速度が最高速度の場合は0が選択され、そうでない場合は、ピッチパターン修正部1404の出力である修正されたアクセント成分値とフレーズ成分値が選択されることになる。

【0100】

選択されたデータは基底ピッチ加算部1407に送られる。基底ピッチ加算部1407にはユーザからの声の高さ指定レベルが入力されており、基底ピッチテーブル1409から該レベルに対応する基底ピッチデータが読み出され、前述のスイッチ1405からの出力値との加算処理が施され、ピッチパタンの時系列データとして合成パラメータ生成部1307に出力される。

【0101】

例えば、声の高さレベルが5段階、レベル0からレベル4まで設定できるシステムの場合、基底ピッチテーブル1409に格納されているデータは男声音の場合、3.0、3.2、3.4、3.6、3.8といった数値、女性音の場合は、4.0、4.2、4.4、4.6、4.8といった数値が良く用いられる。

【0102】

上記の例では、ピッチパターン修正部1404の出力と数値0とをスイッチ1405で切り替える処理を行っているが、無論、発声速度指定が最高レベルの時は、制御要因設定部1401からピッチパターン修正部1404までの処理は不要になる。

【0103】

図8に第2の実施の形態におけるピッチパターン生成処理のフローチャートを示す。ここで図中の記号は以下の通りとする。すなわち、入力文章中に含まれるフレーズ総数を I 、単語総数を J 、第 i 番目のフレーズ成分の大きさを A_{pi} 、第 j 番目のアクセント成分の大きさを A_{aj} 、第 j 番目のアクセント句に対して指定される抑揚制御係数 E_j 、とする。

【0104】

ステップST101からステップST106にかけては、フレーズ成分の大きさ A_{pi} の算出を行う。まずステップST101で、フレーズカウンタ i を0に初期化する。次いでステップST102で発声速度レベルの判定を行い、発声速度が最高速度である場合はステップST104に進み、そうでない場合はステップST103に進む。ステップST104では、第 i 番目のフレーズ成分の大きさ A_{pi} を0に設定してステップST105に進む。一方ステップST103では数量化I類などの統計的手法を用いて第 i 番目のフレーズ成分の大きさ A_{pi} が予測され、ステップST105に進む。ステップST105においては、フレーズカウンタ i を1インクリメントする。次いでステップST106で入力文章中のフレーズ総数 I との比較を行い、フレーズカウンタ i が文内フレーズ総数 I を超えた場合、すなわち全てのフレーズに対する処理が終了した場合にフレーズ成分生成処理を終え、ステップST107に進む。そうでない場合は、ステップST102に戻り次のフレーズに対する処理を前述と同様に繰り返す。

【0105】

ステップST107からステップST113にかけては、アクセント成分の大きさ A_{aj} の算出を行う。まずステップST107で、単語カウンタ j を0に初期化する。次いでステップST108で発声速度レベルの判定を行い、発声速度が最高速度である場合はステップST111に進み、そうでない場合はステップST109に進む。ステップST111では、第 j 番目のアクセント成分の大きさ A_{aj} を0に設定してステップST112に進む。一方ステップST109では数量化I類などの統計的手法を用いて第 j 番目のアクセント成分の大きさ A_{aj} が予測され、ステップST110に進む。ステップST110では、第 j 番目

のアクセント句に対して抑揚修正処理が下式により行われる。

$$A_{aj} = A_{aj} \times E_j \quad \dots (4)$$

【0106】

ここで E_j は、ユーザが指定する抑揚制御レベルに応じてあらかじめ定められている抑揚制御係数であり、先にも説明したように例えば抑揚制御レベルが3段階で与えられ、レベル0が抑揚を1.5倍に、レベル1が抑揚を1.0倍に、レベル2が抑揚を0.5倍にといった場合は以下のようなになる。

レベル0 (抑揚を1.5倍)	$E_j = 1.5$
レベル1 (抑揚を1.0倍)	$E_j = 1.0$
レベル2 (抑揚を0.5倍)	$E_j = 0.5$

【0107】

抑揚修正終了後ステップST112に進む。ステップST112においては、単語カウンタ j を1インクリメントする。次いでステップST113で入力文章中の単語総数 J との比較を行い、単語カウンタ j が文内単語総数 J を超えた場合、すなわち全て単語に対する処理が終了した場合にアクセント成分生成処理を終え、ステップST114に進む。そうでない場合は、ステップST108に戻り次のアクセント句に対する処理を前述と同様に繰り返す。

【0108】

ステップST114では、上記の処理で決定されたフレーズ成分値 A_{pi} とアクセント成分値 A_{aj} 、基底ピッチテーブル1409を参照して得られる基底ピッチ $l_n F_{min}$ とから式(1)によりピッチパターンを生成する。

【0109】

以上詳細に説明したように本発明の第2の実施の形態によれば、発声速度が既定値最大に設定された場合に、ピッチパターンの抑揚成分を0にしてピッチパターン生成を行うため、時間的に速い周期で抑揚が変動することがなくなり、非常に聞き取りにくい合成音となることが解消される。

【0110】

図9は従来技術における発声速度によるピッチパターンの違いの説明図である。上段(a)が通常発声速度の場合であり、下段(b)が最高速度の場合である。

横軸が時間であり、図中点線で示す曲線がフレーズ成分を表わし、実線で示す曲線がアクセント成分に対応している。最高速度が通常速度の2倍だとすると、生成される波形は通常時の約 $1/2$ となる。 $(T_2 = T_1 / 2)$ ピッチパタンの遷移も発声速度に比例して速くなるため、合成音声の抑揚は非常に速い周期での変動となることが図を見ても分かる。しかし実際の発声においては発声速度に応じて、フレーズの結合によるフレーズ境界の消失、アクセント結合によるアクセント句境界の消失といった現象が見られるため図(b)のようにはならない。発声速度が速くなるにつれて、ピッチパタンの変化も相対的に緩やかになることが多い。

【0 1 1 1】

例えば図9の例で言えば2つのフレーズで構成されているが、これが1つのフレーズとして結合するといった現象が確認されている。従来技術においては、この点を考慮に入れておらず、非常に聞きづらい合成音声となっていたが、第2の実施の形態によれば、抑揚成分を0にすることで聞き取り易い合成音声を生成することが可能となる。

【0 1 1 2】

抑揚成分を0にすることで抑揚の全くない、平坦なロボット音声のようになってしまうが、最高速度での合成音出力は通常、読み飛ばしという意味合いで利用される場合がほとんどある。したがって、音声出力されるテキストの内容を把握・理解できれば良い、という程度の使用方法なので、抑揚のない合成音声は使用に耐え得るものである。

【0 1 1 3】

第3の実施の形態

〔構成〕

発明の第3の実施の形態における構成を図面を参照しながら詳細に説明する。本実施の形態が従来技術と異なる点は、文章間に合図音を入れることで文と文との境界を明示する点である。

【0 1 1 4】

図10は、第3の実施の形態におけるパラメータ生成部102の機能ブロック

図であり、この図を用いて説明する。パラメータ生成部 1 0 2 への入力は従来と同じく、テキスト解析部 1 0 1 から出力される中間言語および、ユーザが個別に指定する韻律制御パラメータである。ユーザからの韻律制御指定には、従来技術あるいは第 1、第 2 の実施の形態にはないパラメータとして、合図音指定入力がある。これは後述する、文章間に挿入する合図音の種類を指定するための入力である。

【0 1 1 5】

中間言語解析部 1 7 0 1 には一文毎の中間言語が入力され、以降の韻律生成処理で必要となる音韻系列・フレーズ情報・アクセント情報などといった中間言語解析結果が、それぞれピッチパターン決定部 1 7 0 2、音韻継続時間決定部 1 7 0 3、音韻パワー決定部 1 7 0 4、音声素片決定部 1 7 0 5、声質係数決定部 1 7 0 6 に出力される。

【0 1 1 6】

ピッチパターン決定部 1 7 0 2 には、前述の中間言語解析結果に加えてユーザからの抑揚指定・声の高さ指定・発声速度指定・話者指定の各パラメータが入力され、ピッチパターンが合成パラメータ生成部 1 7 0 8 に出力される。

【0 1 1 7】

音韻継続時間決定部 1 7 0 3 には、前述の中間言語解析結果に加えてユーザからの発声速度指定のパラメータが入力され、それぞれの音韻継続時間・ポーズ長といったデータが合成パラメータ生成部 1 7 0 8 に出力される。

【0 1 1 8】

音韻パワー決定部 1 7 0 4 には、前述の中間言語解析結果に加えてユーザからの声の大きさ指定パラメータが入力され、それぞれの音韻振幅係数が合成パラメータ生成部 1 7 0 8 に出力される。

【0 1 1 9】

音声素片決定部 1 7 0 5 には、前述の中間言語解析結果に加えてユーザからの話者指定パラメータが入力され、波形重畳するための必要な音声素片アドレスが合成パラメータ生成部 1 7 0 8 に出力される。

【0 1 2 0】

声質係数決定部 1 7 0 6 には、前述の中間言語解析結果に加えてユーザからの声質指定パラメータが入力され、声質変換パラメータが合成パラメータ生成部 1 7 0 8 に出力される。

【 0 1 2 1 】

合図音決定部 1 7 0 7 には、ユーザからの発声速度指定・合図音指定パラメータが入力され、合図音の種類および制御用のための合図音制御信号が波形生成部 1 0 3 に出力される。

【 0 1 2 2 】

合成パラメータ生成部 1 7 0 8 は、入力された各韻律パラメータ（前述したピッチパターン、音韻継続時間、ポーズ長、音韻振幅係数、音声素片アドレス、声質変換係数）から、フレーム（通常 8 m s 程度の長さ）を一つの単位とした波形生成用のパラメータに変換し、波形生成部 1 0 3 に出力する。

【 0 1 2 3 】

パラメータ生成部 1 0 2 において、従来技術と比較して異なる点は、合図音決定部 1 7 0 7 が新たな機能ブロックとして存在していることと、その入力パラメータとしてユーザから合図音指定がある点および、波形生成部 1 0 3 の内部構成である。テキスト解析部 1 0 1 においては、従来と同様であるため、その構成に関する説明は省略する。

【 0 1 2 4 】

はじめに合図音決定部 1 7 0 7 の構成について図 1 1 を用いて説明する。図に示すように、合図音決定部 1 7 0 7 は単にスイッチの役割を果たす機能ブロックである。ユーザから指定される発声速度レベルはスイッチ 1 8 0 1 の制御用端子に接続され、同じくユーザから指定される合図音コードがスイッチ 1 8 0 1 の a 端子に接続される。スイッチ 1 8 0 1 の b 端子は常にグランドに接続されている。スイッチ 1 8 0 1 は、発声速度レベルによって、端子 a、端子 b のいずれかに接続されるように構成されている。発声速度が最高レベルの場合はスイッチ 1 8 0 1 を a 端子に接続し、それ以外の場合はスイッチ 1 8 0 1 を b 端子に接続する。すなわちスイッチ 1 8 0 1 は、発声速度が最高レベルの時には合図音コードを、それ以外の時には 0 を出力する構成となっている。スイッチ 1 8 0 1 の出力は、

合図音制御信号として波形生成部 1 0 3 に出力される。

【 0 1 2 5 】

次に波形生成部 1 0 3 の構成について図 1 2 を用いて説明する。第 3 の実施の形態においては、波形生成部 1 0 3 は、素片復号部 1 9 0 1 と振幅制御部 1 9 0 2 と素片加工部 1 9 0 3 と重畳制御部 1 9 0 4 と合図音制御部 1 9 0 5 と D A リングバッファ 1 9 0 6 の各機能ブロック、および合図音辞書 1 9 0 7 とから構成されている。

【 0 1 2 6 】

前述したパラメータ生成部 1 0 2 からの出力は、合成パラメータとして素片復号部 1 9 0 1 に入力される。素片復号部 1 9 0 1 には素片辞書 1 0 5 が接続されており、入力された合成パラメータのうち、素片アドレスを参照ポインタとして素片辞書 1 0 5 から素片データをロードし、必要に応じて復号処理を行い、復号素片データを振幅制御部 1 9 0 2 に出力する。素片辞書 1 0 5 には、音声を合成するための元となる音声素片データが格納されており、記憶容量の節約のために何らかの圧縮処理が施されている場合がある。この時は復号処理を施し、その必要がない非圧縮素片の場合は、単に読み込んでくるだけの処理となる。

【 0 1 2 7 】

振幅制御部 1 9 0 2 には、前述の復号後の音声素片データと合成パラメータとが入力されており、合成パラメータのうち音韻振幅係数によって素片データのパワー制御が行われ、素片加工部 1 9 0 3 に出力される。

【 0 1 2 8 】

素片加工部 1 9 0 3 には、前述の振幅制御された素片データと合成パラメータとが入力されており、合成パラメータのうち声質変換係数によって素片データの伸縮処理が施され、重畳制御部 1 9 0 4 に出力される。

【 0 1 2 9 】

重畳制御部 1 9 0 4 には、前述の伸縮処理が施された素片データと合成パラメータとが入力されており、合成パラメータのうちピッチパターン、音韻継続時間、ポーズ長といったパラメータを用いて素片データの波形重畳処理を施す。重畳制御部 1 9 0 4 で生成される波形は、逐次 D A リングバッファ 1 9 0 6 に出力され

書き込まれる。DAリングバッファ1906に書き込まれたデータは、当該テキスト音声変換システムで設定されている出力サンプリング周期で、図示していないDAコンバータに送られ、合成音がスピーカなどから出力される。

【0130】

波形生成部103には、前述したパラメータ生成部102からの出力として合図音制御信号が合図音制御部1905に入力される。合図音制御部1905にはさらに合図音辞書1907が接続されており、これに格納されているデータを必要に応じて加工してDAリングバッファ1906に出力する。ただし書き込むタイミングは、重畳制御部1904が1文章分の合成波形を出力し終えた後あるいは、合成波形を書き込む前とする。

【0131】

合図音辞書1907には例えば、各種効果音データのPCM(Pulse Code Modulation)データで構築されている構成でも、基準正弦波データが格納された構成でも、どの形態でも構わない。この場合、合図音制御部1905は、前者の辞書構成においては合図音辞書1907からデータを読み出してきて、そのままDAリングバッファ1906に出力し、後者の辞書構成においては合図音辞書1907からデータを読み出し、それを繰り返しつなぎ合わせるなどして出力する。合図音制御部1905に接続されている合図音制御信号が0の場合は、DAリングバッファ1906に出力する処理は行わない。

【0132】

〔動作〕

以上のように構成された第3の実施の形態における動作について図10～図12を用いて詳細に説明する。従来技術と異なる点は、ピッチパターン生成と波形生成に関わる処理であるので、それ以外の処理については省略する。

【0133】

まず、テキスト解析部101で生成された中間言語は、パラメータ生成部102内部の中間言語解析部1701に送られる。中間言語解析部1701では、中間言語上に記述されているフレーズ区切り記号、単語区切り記号、アクセント核を示すアクセント記号、そして音韻記号列から、韻律生成に必要なデータを抽出

して、ピッチパターン決定部1702、音韻継続時間決定部1703、音韻パワー決定部1704、音声素片決定部1705、声質係数決定部1706のそれぞれの機能ブロックへ送る。

【0134】

ピッチパターン決定部1702では、声の高さの遷移であるイントネーションが生成され、音韻継続時間決定部1703では、音韻個々の継続時間のほか、フレーズとフレーズの切れ目あるいは、文と文との切れ目に挿入するポーズ長を決定する。また、音韻パワー決定部1704では、音声波形の振幅値の遷移である音韻パワーが生成され、音声素片決定部1705では合成波形を生成するために必要となる音声素片の、素片辞書105におけるアドレスを決定する。声質係数決定部1706では、素片データを信号処理で加工するためのパラメータの決定が行われる。ユーザから指定される韻律制御指定のうち、抑揚指定および声の高さ指定はピッチパターン決定部1702に、発声速度指定は音韻継続時間決定部1703と合図音決定部1707に、声の大きさ指定は音韻パワー決定部1704に、話者指定はピッチパターン決定部1702と音声素片決定部1705に、声質指定は声質係数決定部1706に、合図音指定は合図音決定部1707に、それぞれ送られている。

【0135】

各機能ブロックのうち、ピッチパターン決定部1702、音韻継続時間決定部1703、音韻パワー決定部1704、音声素片決定部1705、声質係数決定部1706については、従来技術と同様であるのでここでは説明を省略する。

【0136】

第3の実施の形態におけるパラメータ生成部102が従来技術と異なる点は、合図音決定部1707が新たに加えられたことであるので、合図音決定部1707の動作について図11を用いて説明する。図に示すように、合図音決定部1707は単にスイッチの役割を果たす機能ブロックである。スイッチ1801は、ユーザから指定される発声速度レベルによって制御されるような構成を有しており、これにより端子a、端子bのいずれかに接続されるようになっている。制御信号である発声速度レベルが最高速度の時は、スイッチ1801をa端子に接続

し、それ以外の場合はスイッチ1801をb端子に接続する。a端子には、ユーザから指定される合図音コードが入力されており、b端子にはグランド・レベルすなわち0が入力されている。すなわちスイッチ1801は、発声速度が最高レベルの時には合図音コードを、それ以外の時には0を出力する構成となっている。スイッチ1801の出力は、合図音制御信号として波形生成部103に送られる。

【0137】

次に波形生成部103の動作について図12を用いて説明する。パラメータ生成部102内の合成パラメータ生成部1708で生成された合成パラメータは、波形生成部103内の素片復号部1901と振幅制御部1902と素片加工部1903と重畳制御部1904に送られる。

【0138】

素片復号部1901では、合成パラメータのうち、素片アドレスを参照ポイントとして素片辞書105から素片データをロードし、必要に応じて復号処理を行い、復号素片データを振幅制御部1902に送る。素片辞書105には合成波形を生成するための元となる音声素片が格納されており、これをピッチパターンで示される周期で重ね合わせていくことにより音声波形を生成するしくみとなっている。

【0139】

ここで音声素片とは、接続して合成波形を作るための音声の基本単位で、音の種類等に応じて様々なものが用意されている。一般的に、CV、VV、VCV、CVC（C：子音、V：母音）といった音韻連鎖で構成されている場合が多い。上記のように、同じ音韻の素片であっても、前後の音韻環境によって様々な単位で構築されているためデータ容量は膨大となる。そのため通常は、ADPCM（Adaptive Differential PCM）符号化や、周波数パラメータと駆動音源データの対で構成するといった、圧縮技術を施す場合が多い。無論、圧縮を行わずPCMデータとして構築されている場合もある。素片復号部1901によって復元された音声素片データは、振幅制御部1902に送られパワー制御が施される。

【 0 1 4 0 】

振幅制御部 1 9 0 2 には、合成パラメータのうち振幅係数が入力されており、先の音声素片データに乗じられて振幅制御が施される。振幅係数は、ユーザから指定される声の大きさレベル、音韻の種類、呼気段落内での音節位置、該音韻内での位置（立ち上がり区間・定常区間・立ち下がり区間）など、様々な情報から経験的に決定されている。振幅制御された音声素片は、素片加工部 1 9 0 3 に送られる。

【 0 1 4 1 】

素片加工部 1 9 0 3 では、ユーザから指定された声質変換レベルに応じて素片データの伸縮処理（リサンプリング）が施される。声質変換とは、素片辞書 1 0 5 に登録されている素片データに、信号処理等の加工を施すことにより、聴感上、別話者として取り扱えるようにした機能である。一般に、素片データを線形に伸縮する処理を施して実現する場合が多い。伸長処理は、素片データのオーバーサンプリング処理で実現され、太い声となる。逆に縮小処理は、素片データのダウンサンプリング処理で実現され、細い声となる。同一データで別話者を実現するための機能であるため、声質変換処理は上記の手法に限るものではない。また、ユーザからの声質変換指定がない場合は当然のことながら、素片加工部 1 9 0 3 での処理は一切行われない。

【 0 1 4 2 】

以上の処理によって生成された音声素片は、重畳制御部 1 9 0 4 で波形重畳処理が施される。一般的に、ピッチパターンで示されたピッチ周期で素片データをずらしながら重ね合わせて加算するという手法が用いられる。

【 0 1 4 3 】

このようにして生成された合成波形は、逐次 D A リングバッファ 1 9 0 6 に書き込まれ、当該テキスト音声変換システムで設定されている出力サンプリング周期で、図示していない D A コンバータに送られ、合成音がスピーカなどから出力される。

【 0 1 4 4 】

波形生成部 1 0 3 にはさらに、パラメータ生成部 1 0 2 内の合図音決定部 1 7

07から送られる合図音制御信号が入力されている。合図音制御信号は、合図音制御部1905を介して合図音辞書1907に登録されているデータをDAリングバッファ1906に書き込むための信号である。合図音制御信号が0の場合、すなわち前述したように、ユーザから指定される発声速度が最高速度レベルではない時は、合図音制御部1905は一切の処理を行わない。0以外の場合、すなわち前述したように、ユーザから指定される発声速度が最高速度レベルの時は、合図音制御信号を合図音の種類とみなして合図音辞書1907からのデータロードを行う。

【0145】

例えば、合図音の種類を3種類設ける。合図音辞書1907には、例えば、500Hzの正弦波データ、1KHzの正弦波データ、2KHzの正弦波データがそれぞれ1周期分格納されており、それらを複数回繰り返し接続することにより「ピッ」という合図音を生成することとする。合図音制御信号の取り得る値は、0、1、2、3の4種類となり、0の時は一切の処理を行わず、1の時は合図音辞書1907から500Hzの正弦波データを読み出してきて、それらを既定回数繰り返し接続してDAリングバッファ1906に書き込む。1の時は合図音辞書1907から1KHzの正弦波データを読み出してきて、それらを既定回数繰り返し接続してDAリングバッファ1906に書き込む。2の時は合図音辞書1907から2KHzの正弦波データを読み出してきて、それらを既定回数繰り返し接続してDAリングバッファ1906に書き込む。ただし書き込むタイミングは、重畳制御部1904が1文章分の合成波形を出力し終えた後あるいは、合成波形を書き込む前である。したがって、合図音が出力されるのは文章間ということになる。出力される正弦波データは、100ms～200ms程度が適当と思われる。

【0146】

また、正弦波データではなく、出力されるべき合図音を直接PCMデータとして合図音辞書1907に格納しておくという構成でも構わない。この場合、合図音辞書1907からデータを読み出してきて、そのままDAリングバッファ1906に出力する処理が施されることになる。

【 0 1 4 7 】

以上詳細に説明したように、第 3 の実施の形態によれば、発声速度が既定値最大に設定された場合に、文章と文章の間に合図音を挿入する機能を有しているため、早聞き機能有効時での従来技術での問題点である、文境界が把握しにくく、読上げテキストの内容理解が困難であるといったことが解消される。

【 0 1 4 8 】

例えば、以下の文言をテキスト合成する場合を考える。

「出席予定者：開発部 山田部長。企画室 斉藤室長。営業 1 部 渡辺部長。」
処理単位、すなわち 1 文章の区切り記号は句点「。」とすると、上記の文言は以下の 3 文章からなる。

(1) 「出席予定者：開発部 山田部長。」

(2) 「企画室 斉藤室長。」

(3) 「営業 1 部 渡辺部長。」

従来技術によれば、発声速度が速くなるとそれぞれの文終端におけるポーズ長も短くなるため、文章 (1) の最後の「山田部長」という合成音声と、文章 (2) の先頭の「企画室」という合成音声がほぼ連続して出力されるため、「山田部長」＝「企画室」というような誤った認識を受ける場合も発生する。

【 0 1 4 9 】

しかしながら、第 3 の実施の形態によれば、「山田部長」という合成音声と、「企画室」という合成音声の間に、例えば「ピッ」という合図音が挿入されるため、上記のような誤認識は発生しない。

【 0 1 5 0 】

第 4 の実施の形態

[構成]

本発明の第 4 の実施の形態における構成を図 1 3 を参照しながら詳細に説明する。この実施の形態が従来技術と異なる点は、早聞き機能有効時の音韻継続時間の伸縮率決定の際に、現在処理中のテキストが文内における先頭単語あるいは先頭フレーズであるかを判定して、その結果により伸縮係数を決定する点である。したがって、従来と異なる音韻継続時間決定部についてのみ説明し、それ以外の

機能ブロックすなわち、テキスト解析部、波形生成部、音韻継続時間決定部以外のパラメータ生成部内部モジュールについては説明を省略する。

【0151】

音韻継続時間決定部203への入力は従来と同じく、中間言語解析部201からの音韻・韻律情報を含んだ解析結果および、ユーザからの指定される発声速度レベルである。1文章に対する中間言語解析結果は制御要因設定部2001と単語カウンタ2005とに接続されている。制御要因設定部2001では、音韻継続時間決定のために必要な制御要因パラメータの解析が行われ、その出力が継続時間推定部2002に接続される。継続時間の決定には数量化I類等の統計的手法を用いており、例えば、音韻長は通常、目標となる音韻の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置などにより予測され、ポーズ長は、前後隣接するフレーズのモーラ総数などといった情報から予測が行われる場合が多い。制御要因設定部2001はこれら予測に必要な情報の抽出を行っている。

【0152】

継続時間推定部2002には、継続時間予測テーブル2004が接続されており、これを用いて継続時間の予測が行われ、継続時間修正部2003に出力される。継続時間予測テーブル2004は、大量の自然発声データを基に数量化I類などの統計的手法を用いて予め学習されたデータである。

【0153】

一方、単語カウンタ2005では、現在解析中の音韻が、文章内における先頭単語あるいは先頭フレーズに含まれているのか、そうでないのかの判定を行い、その結果を伸縮係数決定部2006に出力する。

【0154】

伸縮係数決定部2006にはさらに、ユーザから指定される発声速度レベルが入力されており、現在処理中の音韻に対する音韻継続時間長の修正係数を決定する機能を有しており、これを継続時間修正部2003に接続している。

【0155】

継続時間修正部2003では、継続時間推定部2002で予測された音韻継続時間に対して、伸縮係数決定部2006で決定された伸縮係数を乗じることによ

り、音韻継続時間の修正を行い合成パラメータ生成部に出力する。

【0156】

〔動作〕

以上のように構成された本発明の第4の実施の形態における動作について図13～図14を用いて詳細に説明する。従来技術と異なる点は、音韻継続時間決定に関わる処理であるので、それ以外の処理については省略する。

【0157】

中間言語解析部201から1文章に対応する解析結果が制御要因設定部2001と単語カウンタ2005に入力される。制御要因設定部2001では、音韻継続時間（子音長・母音長・閉鎖区間長）、ポーズ長を決定するために必要な制御要因の設定を行う。音韻継続時間の決定に必要なデータとは、例えば、目標となる音韻の種別、対象音節の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置といった情報である。一方、ポーズ長決定に必要なデータとは、前後隣接するフレーズのモーラ総数といった情報である。これらの継続時間長を決定するために継続時間予測テーブル2004が使用される。

【0158】

継続時間予測テーブル2004は、自然発声データを基に数量化I類などの統計的手法を用いて予め学習したテーブルである。継続時間推定部2002は、このテーブルを参照しながら音韻継続時間、ポーズ長の予測を行う。継続時間推定部2002で算出される個々の音韻継続時間長は、通常発声速度の場合のものである。これらは、継続時間修正部2003において、ユーザから指定された発声速度に応じて修正が施される構成となっている。通常、発声速度指定は、5～10段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。

【0159】

一方、単語カウンタ2005にも、中間言語解析部201から1文章に対応する解析結果が入力されており、現在解析中の音韻が、文章内における先頭単語あるいは先頭フレーズに含まれているのか、そうでないのかの判定が行われる。本

実施の形態では、文章内における先頭単語であるか否かの判定を行う機能として説明を行う。単語カウンタ2005から送られる判定結果は、該音韻が文内先頭単語に含まれている場合にTRUE、そうでない場合にFALSEを出力することとする。単語カウンタ2005での判定結果は伸縮係数決定部2006に送られる。

【0160】

伸縮係数決定部2006には前述の単語カウンタ2005からの判定結果に加えて、ユーザから指定される発声速度レベルが入力されており、これら2つのパラメータから該音韻の伸縮係数の算出を行う。例えば、発声速度レベルが5段階に制御され、発声速度が遅い方からレベル0、レベル1、レベル2、レベル3、レベル4まで指定可能だとする。それぞれのレベル n に対応した定数 T_n を次のように定める。すなわち、

$T_0 = 2.0$ 、 $T_1 = 1.5$ 、 $T_2 = 1.0$ 、 $T_3 = 0.75$ 、 $T_4 = 0.5$ とする。通常発声速度はレベル2となり、早聞き機能が有効とされると発声速度はレベル4に設定されることになる。単語カウンタ2005からの信号がTRUEの場合、発声速度レベルが0～3まで範囲であれば上記 T_n をそのまま継続時間修正部2003に出力する。発声速度レベルが4であれば、通常発声時の T_2 の数値を出力する。単語カウンタ2005からの信号がFALSEの場合は、発声速度レベルに関わらず上記 T_n をそのまま継続時間修正部2003に出力する。

【0161】

継続時間修正部2003では、継続時間推定部2002から送られる音韻継続時間長に対して、伸縮係数決定部2006からの伸縮係数を乗じて修正を施す。ただし修正を行うのは通常、母音長のみである。発声速度レベルに応じた修正が施された音韻継続時間は合成パラメータ生成部へ送られる。

【0162】

さらに詳細に説明するために図14に継続時間決定処理のフローチャートを示す。ここで図中の記号は以下の通りとする。すなわち、入力文章中に含まれる単語総数を I 、第 i 番目の単語を構成する音韻に対する継続時間修正係数を TC_i

、ユーザから指定される発声速度レベルを lev (ただし範囲は $0 \sim 4$ までの 5 段階とし、数値が多いほど速度が速いこととする)、発声速度がレベル n の時の伸縮係数を $T(n)$ 、第 i 番目の単語の第 j 番目の母音長を T_{ij} 、単語を構成する音節数はそれぞれの単語によって変わるがここでは簡単化のために一律 J とする。

【0163】

まずステップ $ST201$ で単語数カウンタ i を 0 に初期化する。次いでステップ $ST202$ で単語数と発声速度レベルの判定が行われる。現在処理中の単語数カウンタが 0 でかつ、発声速度レベルが 4 の時、これはすなわち、現在処理している音節が文内先頭単語に属しており、かつ発声速度が最高レベルの時であるが、この時はステップ $ST204$ に進み、そうでないときはステップ $ST203$ に進む。ステップ $ST204$ では発声速度レベル 2 の値が修正係数として選択され、ステップ $ST205$ に進む。すなわち、

$$TC_i = T(2) \quad \dots (5)$$

となる。

【0164】

ステップ $ST203$ では、ユーザから指定されたレベル通りの修正係数が選択され、ステップ $ST205$ に進む。すなわち、

$$TC_i = T(lev) \quad \dots (6)$$

となる。

【0165】

ステップ $ST205$ では、音節カウンタ j が 0 に初期化されステップ $ST206$ に進む。ステップ $ST206$ では第 i 番目の単語の第 j 番目の母音の継続時間 T_{ij} が、先に求められた修正係数 TC_i によって下式を用いて行われる。

$$T_{ij} = T_{ij} \times TC_i \quad \dots (7)$$

【0166】

次いでステップ $ST207$ で音節カウンタ j が 1 インクリメントされステップ $ST208$ に進む。ステップ $ST208$ では、音節カウンタ j と該単語の音節総数 J との比較を行い、音節カウンタ j が音節総数 J を超えた場合、すなわち該単

語の全ての音節に対する処理が終了した場合にステップST209に進む。そうでない場合は、ステップST206に戻り次の音節に対する処理を前述と同様に繰り返す。

【0167】

ステップST209では単語数カウンタ*i*が1インクリメントされ、次のステップST210に進む。

【0168】

ステップST210では、単語数カウンタ*i*と単語総数*I*との比較を行い、単語数カウンタ*i*が単語総数*I*を超えた場合、すなわち入力文章中の全て単語に対する処理が終了した場合は処理を終了し、そうでない場合は、ステップST202に戻り次の単語に対する処理を前述と同様に繰り返す。

【0169】

上記の処理により、ユーザから指定される発声速度レベルが最高速度となっても、文章先頭単語だけは通常の発声速度での合成音が生成されることになる。

【0170】

以上詳細に説明したように、第4の実施の形態によれば、発声速度が既定値最大に設定された場合に、文先頭の単語に対して音韻継続時間制御を通常の発声速度として処理するため、ユーザが早聞き機能解除のタイミングを計りやすいという効果がある。例えば、ソフトウェア仕様書などのマニュアル類には、「第3章」あるいは「4. 1. 3」などの項目番号が付与されている場合がほとんどある。こういったマニュアル類をテキスト音声変換で読上げを行う際に、第3章から聞きたい、あるいは4. 1. 3節から聞きたいといった場合に、従来技術においては、早聞き機能を有効にした後ユーザが、高速で出力される合成音声の中から「ダイサンショー」あるいは「ヨンテンイッテンサン」といったキーワードを聞き分け、早聞き機能を解除するといった面倒な操作が必要であった。第4の実施の形態によれば、ユーザに負担をかけずに早聞き機能の有効化・無効化を実現することが可能となる。

【0171】

尚、本発明は前述の実施の形態に限定されるものではなく、本発明の趣旨に基

づいて種々変形させることが可能である。例えば、第 1 の実施の形態において、発声速度が既定値最大に設定された場合に、テキスト音声変換処理の中で演算負荷が大きい機能ブロックを簡略化あるいは、無効にする処理を施しているが、この処理は最大発声速度に限らない。つまり、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、高負荷処理として数量化 I 類による韻律パラメータの予測処理、声質変換のための素片データ加工処理を挙げているが、これに限るものではない。他に高負荷処理機能（例えばエコーや高域強調などの音響処理など）を有している場合は当然のことながら、これを無効化あるいは簡略化といった処理形態にすることが望ましい。また、声質変換処理として波形そのものを線形伸縮しているが、非線形伸縮でも、あるいは周波数パラメータに対して規定の変換関数に通して変形するといった方法でも構わない。また、音韻継続時間決定規則、ピッチパターン決定規則を挙げているが、本発明では演算量が少なく済み、処理時間の短縮が図れるモードを有する構成にすること目的としているため、規則化手順は上記に限られるものではない。逆に、通常発声速度の時には、統計的手法を用いた韻律パラメータの予測を行っているが、規則化手順よりも演算負荷がかかる処理であればこれに限るものではない。また、その予測に用いる制御要因を幾つか挙げているがこれはあくまでも一例である。

【 0 1 7 2 】

第 2 の実施の形態において、発声速度が既定値最大に設定された場合に、ピッチパタンの抑揚成分を 0 にしてピッチパターン生成を行っているが、この処理は最大発声速度に限らない。即ち、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、抑揚成分を完全に 0 にしているが、通常時に比べて抑揚成分を弱めるといった方法でも構わない。例えば、発声速度が既定値最大に設定された時は、抑揚指定レベルを強制的に最低レベルに設定し、ピッチパターン修正部において抑揚成分を縮小するといった構成でも構わない。ただこの時の抑揚指定レベルは、高速合成時においても聞き易いイントネーションとなる必要がある。また、ピッチパタンのアクセント成分、フレーズ成分を数量化 I 類によって決定しているが規則によって決定しても無論構わない。また、予測を行う際にその制御要因を幾つか挙げているがこれはあくまでも一例である。

【0173】

第3の実施の形態において、発声速度が既定値最大に設定された場合に、文章と文章の間に合図音を挿入しているが、この処理は最大発声速度に限らない。即ち、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、実施例では基準正弦波の繰り返しにより合図音を生成しているが、ユーザの注意を引けるものであればこれに限らない。録音された効果音をそのまま出力する構成でも構わない。無論、実施例で示したような合図音辞書を持たずに、内部回路あるいはプログラムでその都度生成するような構成でも構わない。またこの実施の形態では1文の合成波形直後に合図音を挿入する構成となっているが、逆に合成波形直前でも構わない。発声速度が既定値最大に設定された時に、ユーザに対して文章境界が明示できればそれでよい。また、この実施の形態ではパラメータ生成部に合図音の種類を指定するための入力が存在するが、ハードウェア規模、ソフトウェア規模の制限などから、これを省略してもよい。しかしながら、ユーザの好みによって合図音を変えることのできる構成の方が好ましい。

【0174】

第4の実施の形態において、発声速度が既定値最大に設定された場合に、文先頭の単語に対して音韻継続時間制御を通常（デフォルト）の発声速度として処理しているが、この処理は最大発声速度に限らない。即ち、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、通常発声速度で処理する単位を文先頭の1単語としているが、先頭2単語あるいは先頭フレーズという構成でも構わない。また、通常の発声速度ではなく、レベルを1段階落とすといった方法も十分考えられる。

【0175】

【発明の効果】

以上詳細に説明したように、請求項1に係る発明によれば、入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片

辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、前記パラメータ生成手段は、音韻継続時間を予め経験的に求めた継続時間規則テーブルと、音韻継続時間を統計的手法を用いて予測した継続時間予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記継続時間規則テーブルを用い、閾値を超えていない時には前記継続時間予測テーブルを用いて音韻継続時間の決定を行う音韻継続時間決定手段を有する構成としたことにより、また、請求項3に係る発明によれば、前記パラメータ生成手段は、アクセント成分及びフレーズ成分を決定するために必要となるデータを、予め経験的に求めた規則テーブルと、統計的手法を用いて予測した予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記規則テーブルを用い、閾値を超えていない時には前記予測テーブルを用いてアクセント成分及びフレーズ成分を決定することによりピッチパターンを決定するピッチパターン決定手段を有する構成としたことにより、更に、請求項5に係る発明によれば、前記パラメータ生成手段は、前記音声素片を変形させて声質を切り換えるための声質変換係数テーブルを備え、ユーザから指定される発声速度が閾値を超えたときには、声質が変化しないような係数を前記声質変換係数テーブルから選択する声質係数決定手段を有する構成としたので、発声速度が既定値最大に設定された場合に、テキスト音声変換処理の中で演算負荷が大きい機能ブロックを簡略化あるいは、無効にする処理を施しているため、高負荷による音切れが発生する機会を減少させ、聞き易い合成音声を生成することが可能となる。

【0176】

また、請求項7に係る発明によれば、前記パラメータ生成手段は、ユーザが指定した抑揚レベルに応じて修正したピッチパターンを出力するするピッチパターン修正手段と、ユーザが指定した発声速度に応じて前記修正したピッチパターンを基底ピッチに加算するか否かを選択する切り換え手段とを有し、前記発声速度が所定の閾値を超えた場合には前記基底ピッチを変更しないように前記切り換え手段を制御する構成としたので、発声速度が既定値最大に設定された場合に、ピッチパ

タンの抑揚成分を0にしてピッチパターン生成を行うため、時間的に速い周期で抑揚が変動することがなくなり、非常に聞き取りにくい合成音となることが解消される。

【0177】

また、請求項10に係る発明によれば、前記波形生成手段は、文章の切れ目であることを示す合図音を文章間に挿入する合図音生成手段を備え、ユーザから指定される発声速度が閾値を超えた時には前記合図音を文章間に挿入する構成としたので、発声速度が既定値最大に設定された場合に、文章と文章の間に合図音を挿入する機能を有しているため、早聞き機能有効時での従来技術での問題点である、文境界が把握しにくく、読上げテキストの内容理解が困難であるといったことが解消される。

【0178】

更に、請求項12に係る発明によれば、前記パラメータ生成手段は、ユーザから指定される発声速度が閾値を超えた時には、少なくとも文章の先頭単語の発声速度を通常の発声速度に戻す処理を行う音韻継続時間決定手段を有する構成としたので、発声速度が既定値最大に設定された場合に、文先頭の単語に対して音韻継続時間制御を通常の発声速度として処理するため、ユーザが早聞き機能解除のタイミングを計りやすいという効果がある。

【図面の簡単な説明】

【図1】

本発明の第1の実施の形態におけるパラメータ生成部の機能ブロック図である。

【図2】

本発明の第1の実施の形態におけるピッチパターン決定部の機能ブロック図である。

【図3】

本発明の第1の実施の形態における音韻継続時間決定部の機能ブロック図である。

【図4】

本発明の第 1 の実施の形態における声質係数決定部の機能ブロック図である。

【図 5】

声質変換のためのデータのリサンプリング周期の説明図である。

【図 6】

本発明の第 2 の実施の形態におけるパラメータ生成部の機能ブロック図である。

【図 7】

本発明の第 2 の実施の形態におけるピッチパターン決定部の機能ブロック図である。

【図 8】

本発明の第 2 の実施の形態におけるピッチパターン生成フローチャートである。

【図 9】

発声速度によるピッチパターンの違いの説明図である。

【図 1 0】

本発明の第 3 の実施の形態におけるパラメータ生成部の機能ブロック図である。

【図 1 1】

本発明の第 3 の実施の形態における合図音決定部の機能ブロック図である。

【図 1 2】

本発明の第 3 の実施の形態における波形生成部の機能ブロック図である。

【図 1 3】

本発明の第 4 の実施の形態における音韻継続時間決定部の機能ブロック図である。

【図 1 4】

本発明の第 4 の実施の形態における継続時間決定フローチャートである。

【図 1 5】

一般的なテキスト音声変換処理の機能ブロック図である。

【図 1 6】

従来技術によるパラメータ生成部の機能ブロック図である。

【図 1 7】

従来技術による波形生成部の機能ブロック図である。

【図 1 8】

ピッチパターン生成過程モデルの説明図である。

【図 1 9】

従来技術によるピッチパターン決定部の機能ブロック図である。

【図 2 0】

従来技術による音韻継続時間決定部の機能ブロック図である。

【図 2 1】

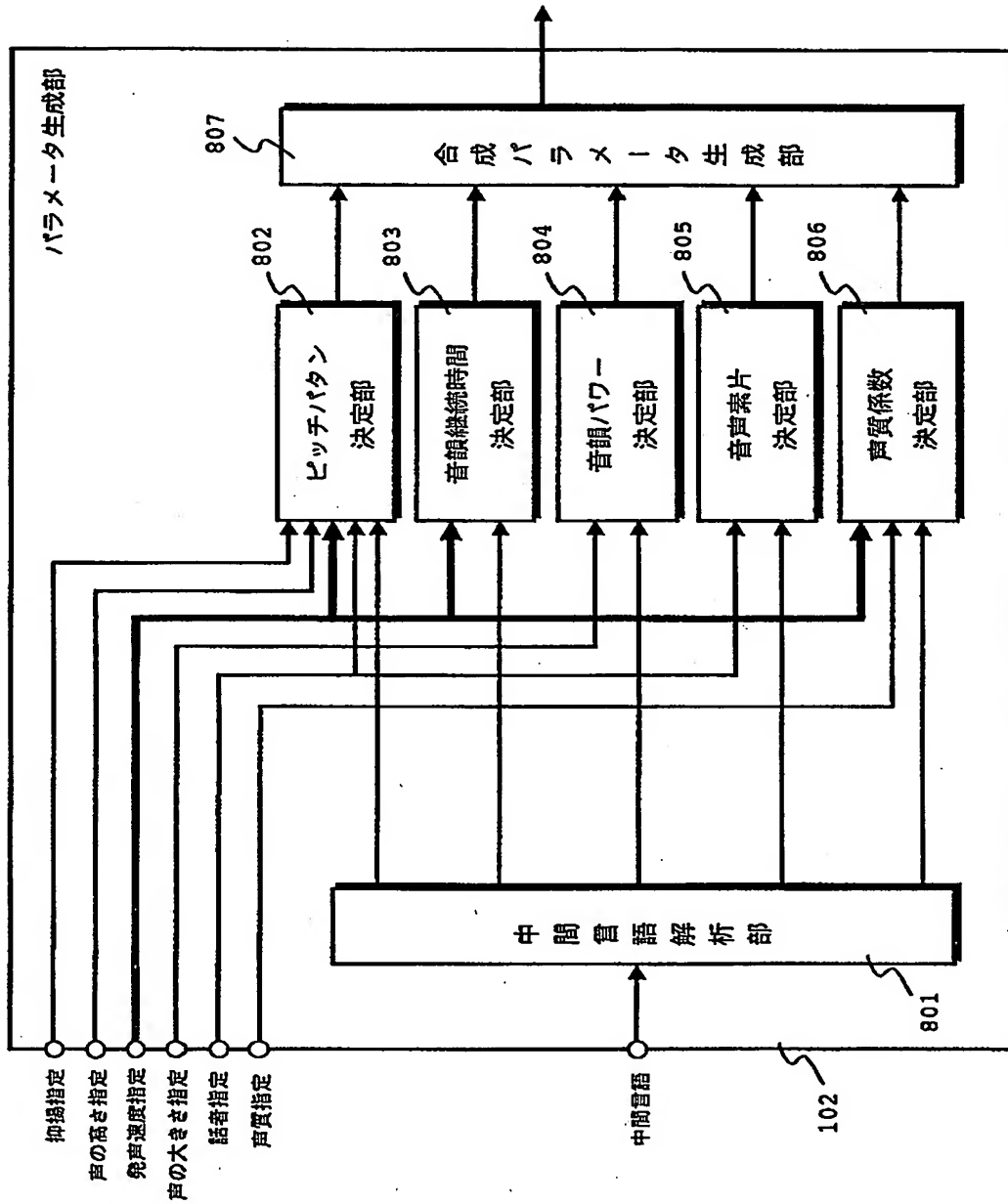
発声速度の違いによる波形伸縮の説明図である。

【符号の説明】

1 0 1	テキスト解析部
1 0 2	パラメータ生成部
1 0 3	波形生成部
1 0 4	単語辞書
1 0 5	素片辞書
8 0 1, 1 3 0 1, 1 7 0 1,	中間言語解析部
8 0 2, 1 3 0 2, 1 7 0 2,	ピッチパターン決定部
8 0 3, 1 3 0 3, 1 7 0 3	音韻継続時間決定部
8 0 4, 1 3 0 4, 1 7 0 4	音韻パワー決定部
8 0 5, 1 3 0 5, 1 7 0 5	音声素片決定部
8 0 6, 1 3 0 6, 1 7 0 6	声質係数決定部
1 7 0 7	合図音決定部
8 0 7, 1 3 0 7, 1 7 0 8	合成パラメータ生成部

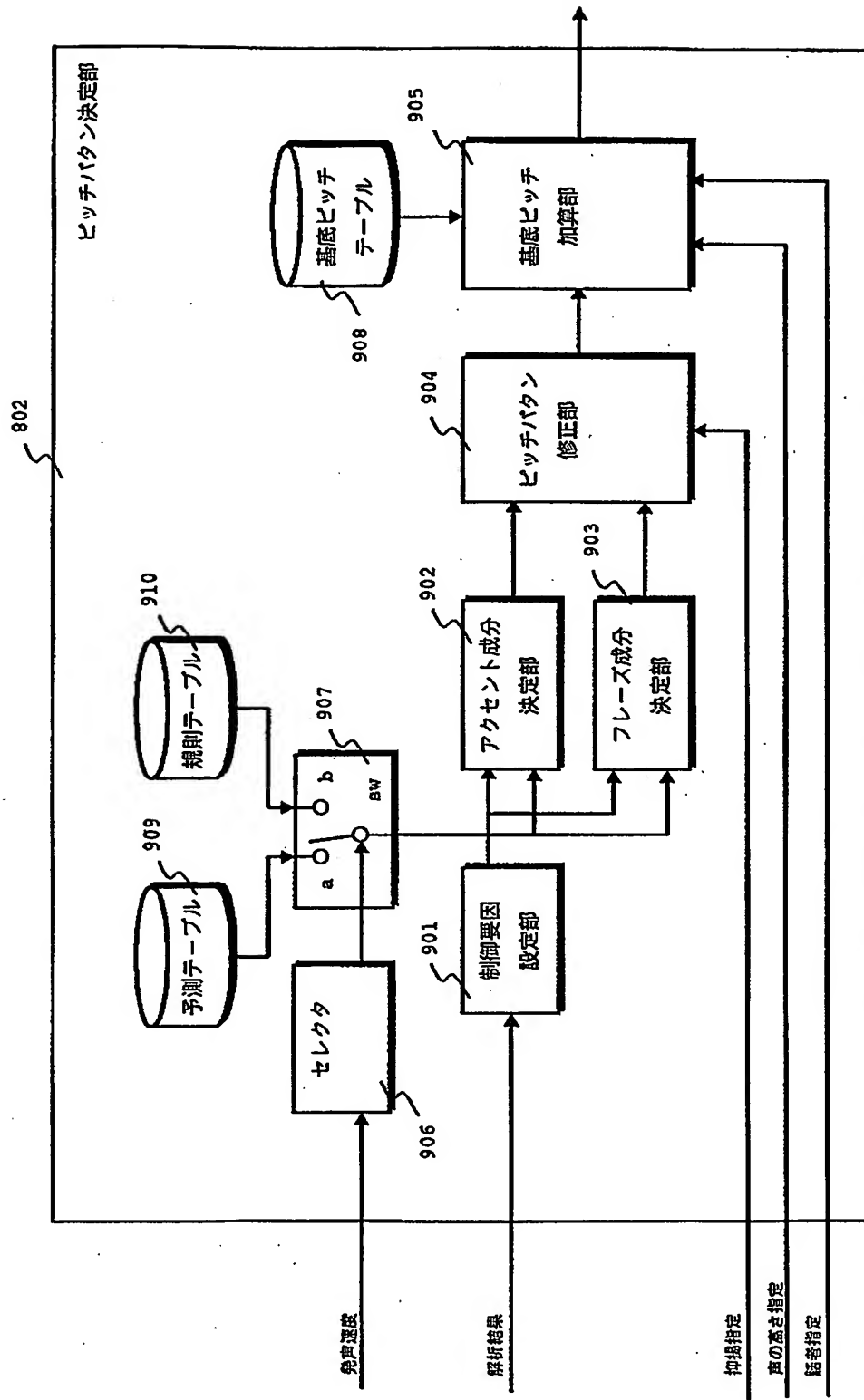
【書類名】 図面

【図 1】



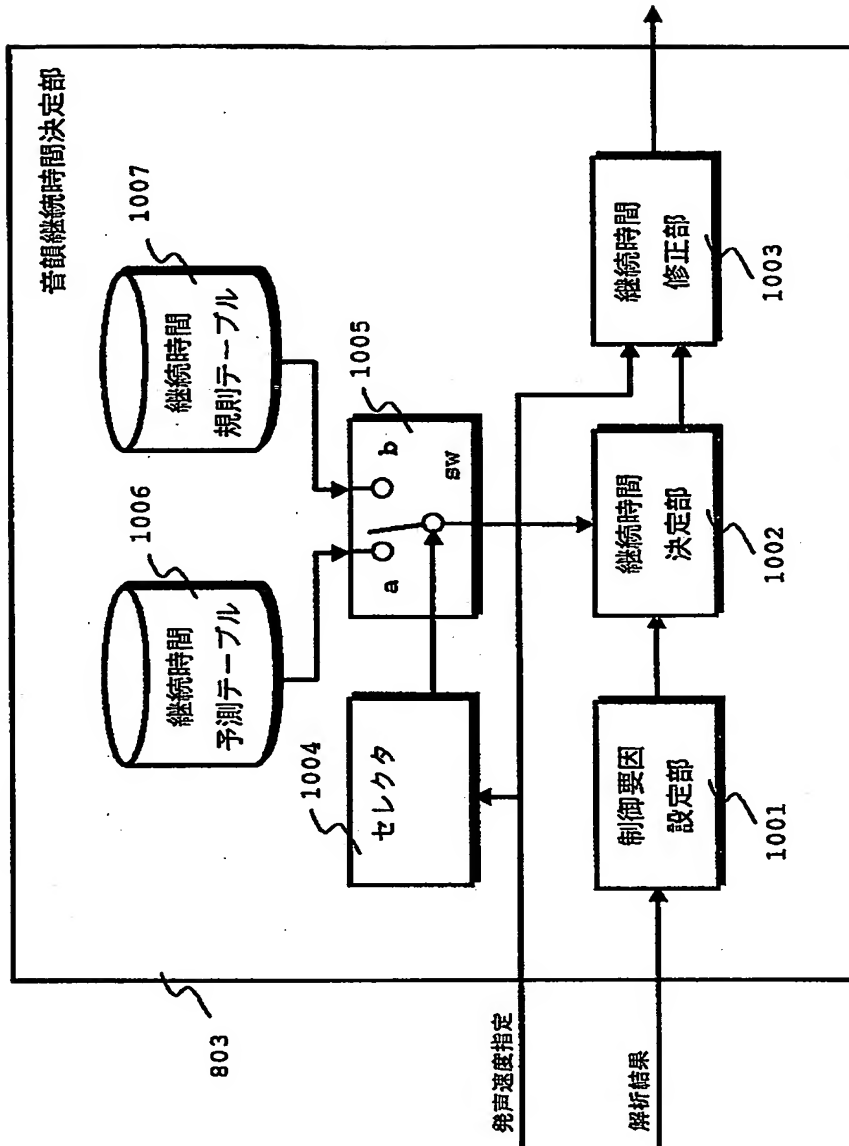
本発明の第1の実施の形態におけるパラメータ生成部の機能ブロック図

【図2】



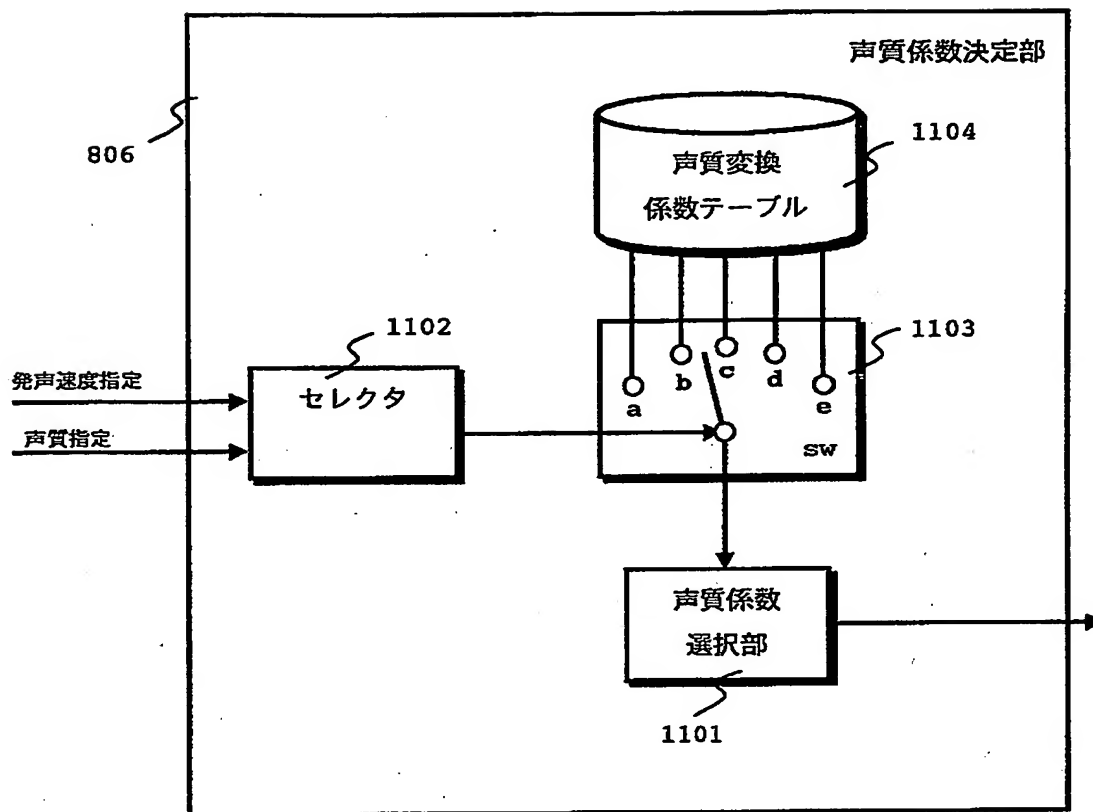
本発明の第1の実施の形態におけるピッチパターン決定部の機能ブロック図

【図 3】



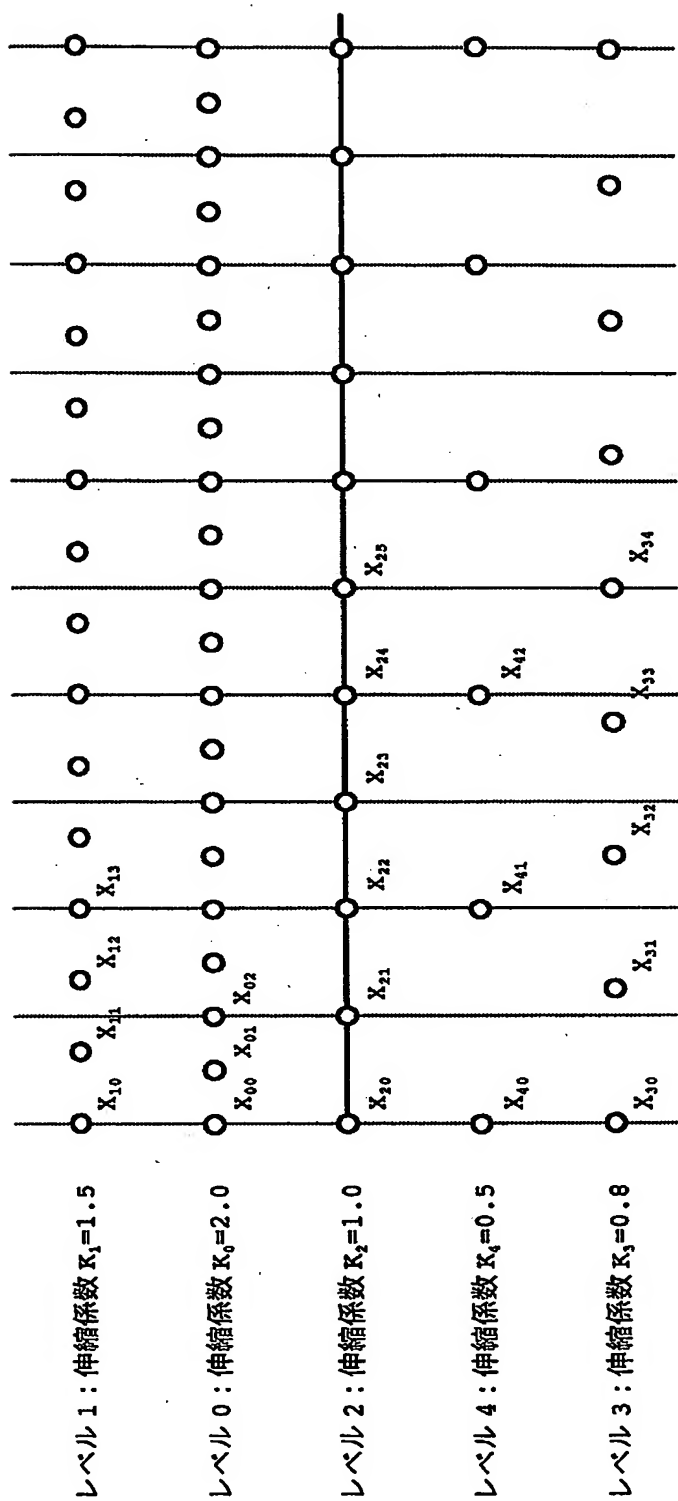
本発明の第 1 の実施の形態における音韻継続時間決定部の機能ブロック図

【図 4】



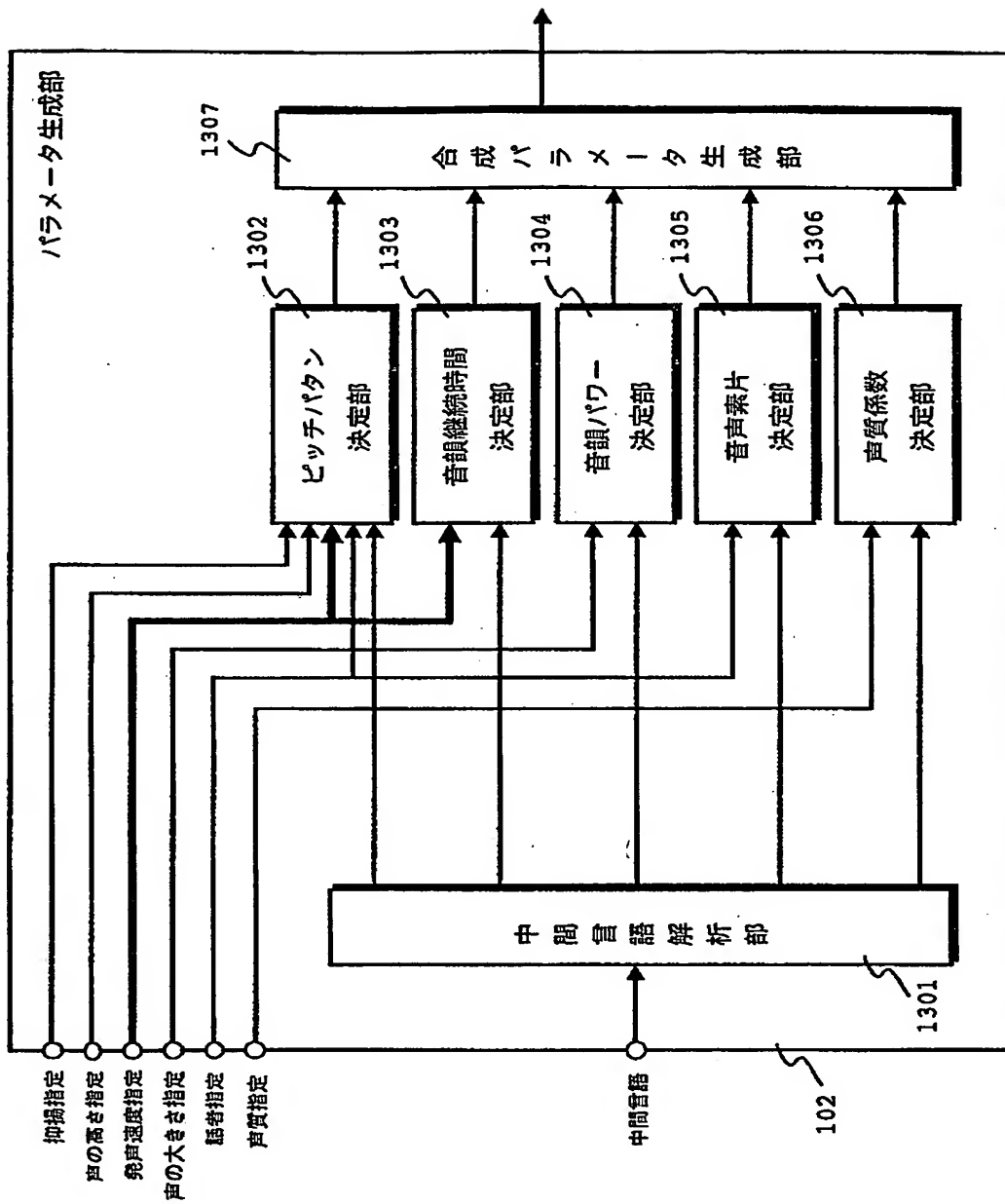
本発明の第 1 の実施の形態における声質係数決定部の機能ブロック図

【図 5】



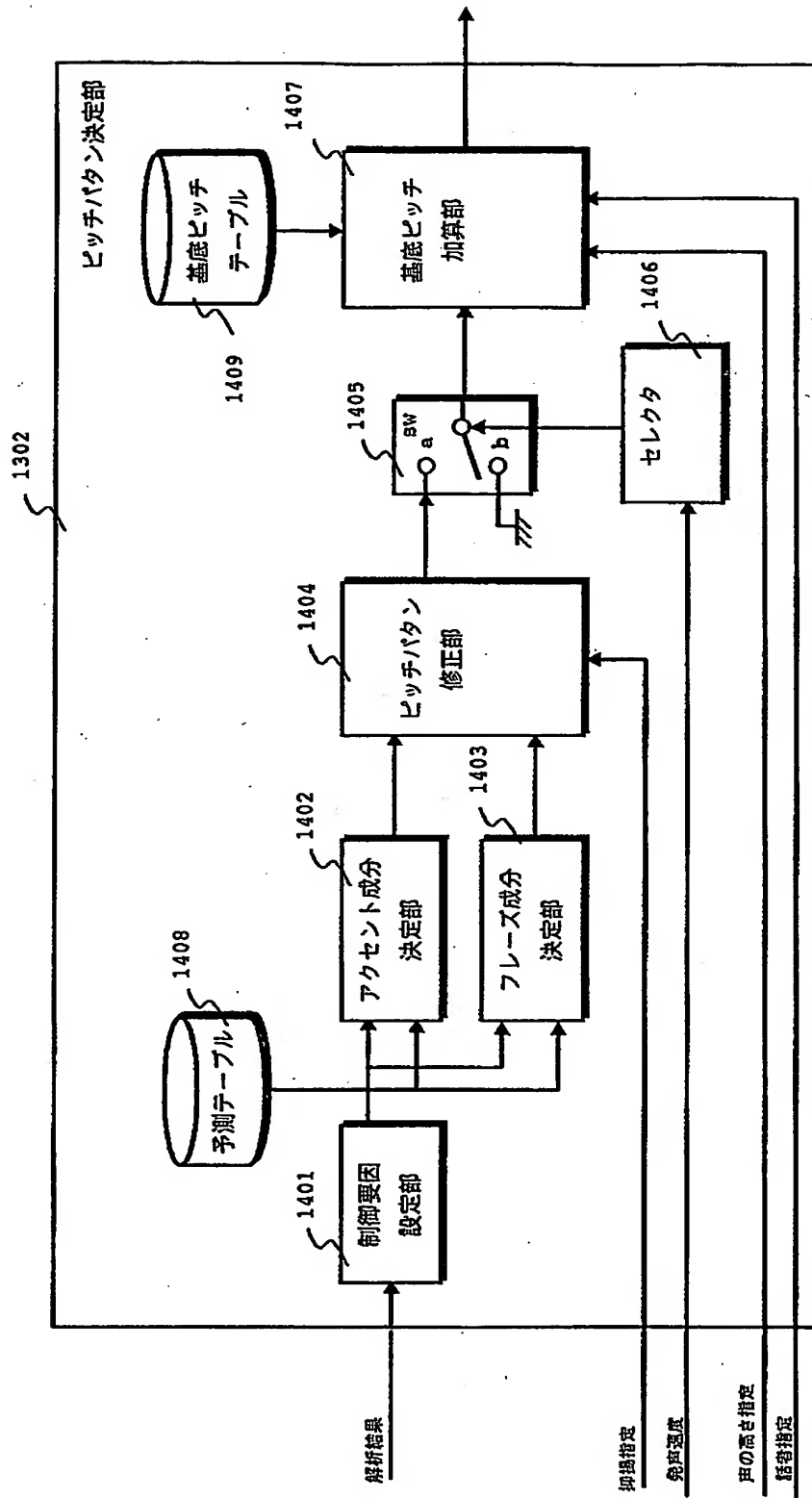
声質変換のためのデータのリスampling周期の説明図

【図 6】



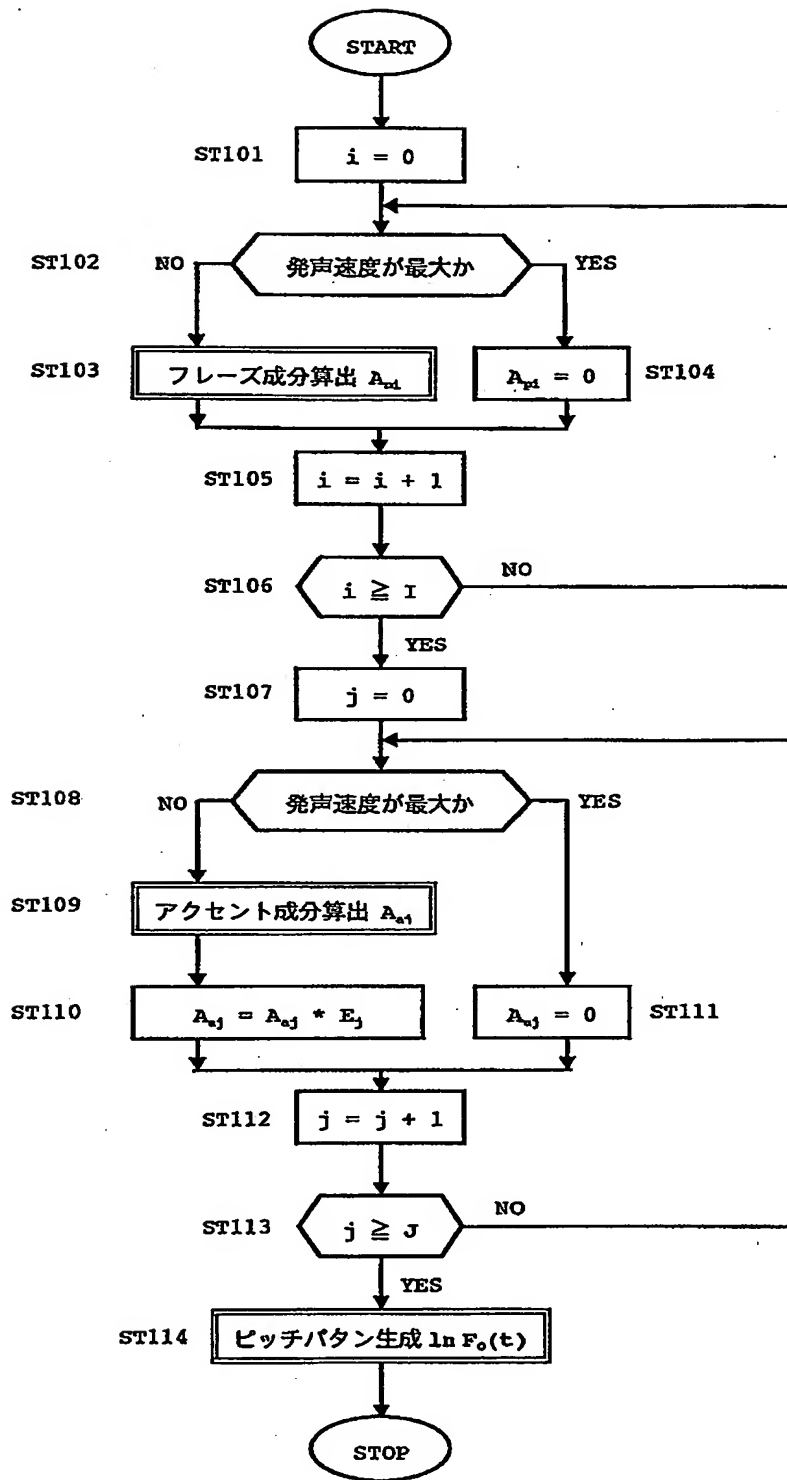
本発明の第 2 の実施の形態におけるパラメータ生成部の機能ブロック図

【図 7】



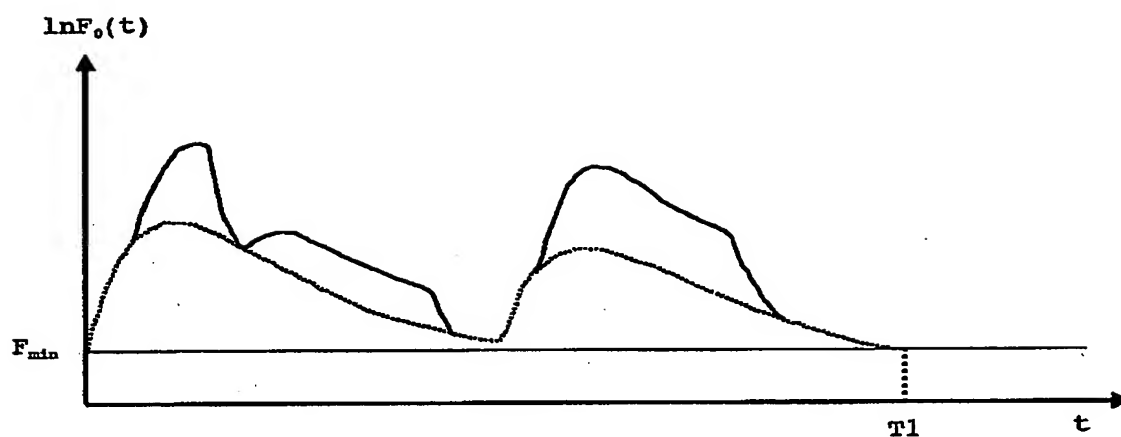
本発明の第2の実施の形態におけるピッチパタン決定部の機能ブロック図

【図 8】

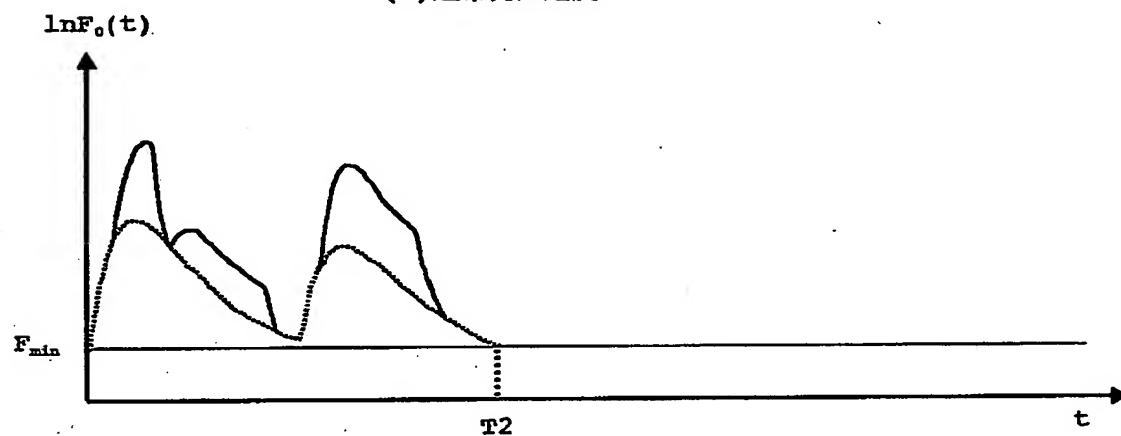


本発明の第2の実施の形態におけるピッチバタン生成フローチャ

【図9】



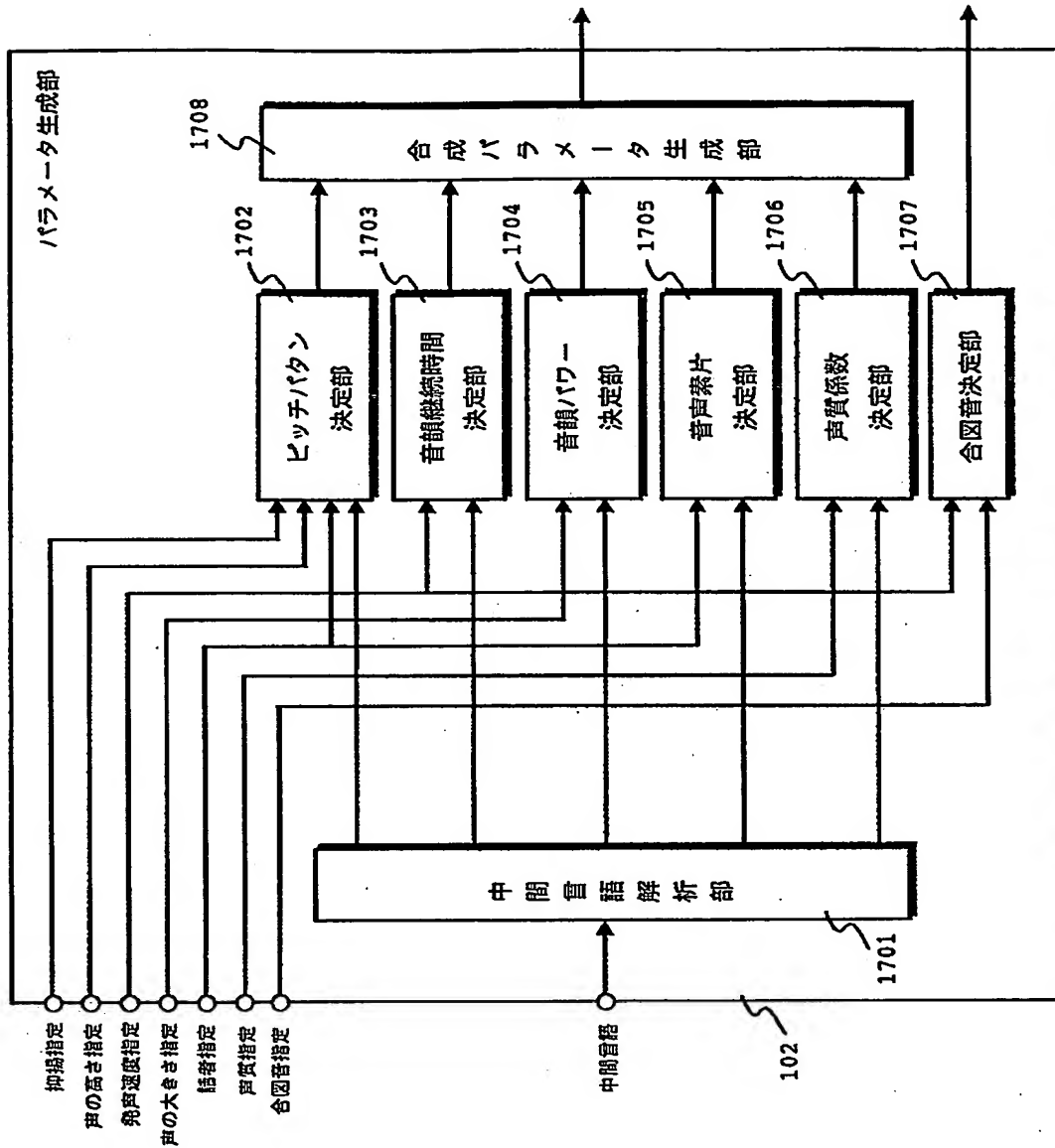
(a)通常発声速度



(b)最高速度

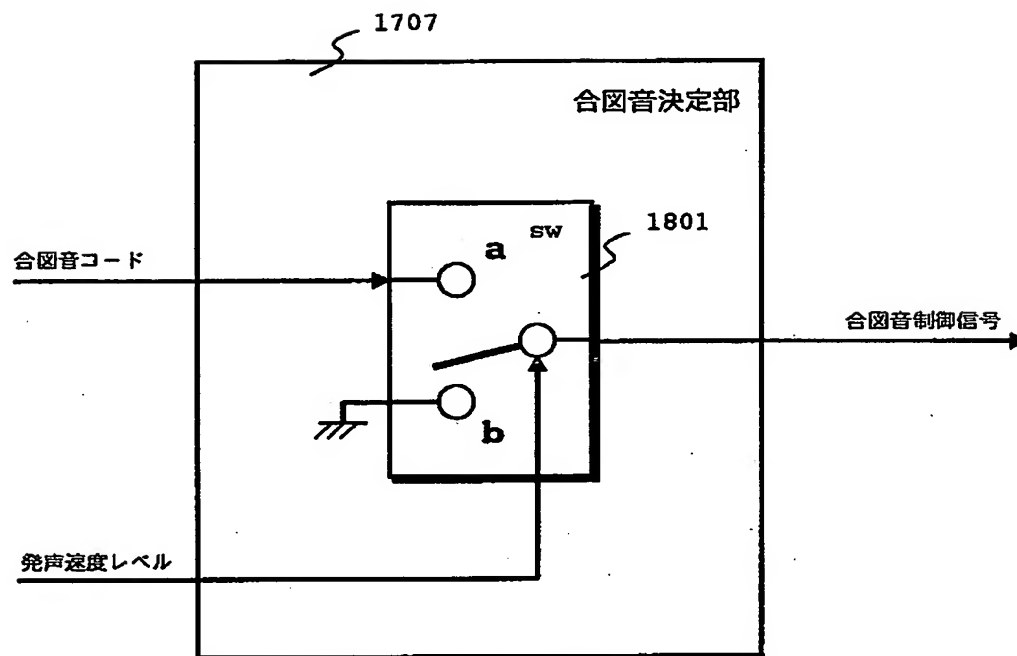
発声速度によるピッチパタンの違いの説明図

【図 10】



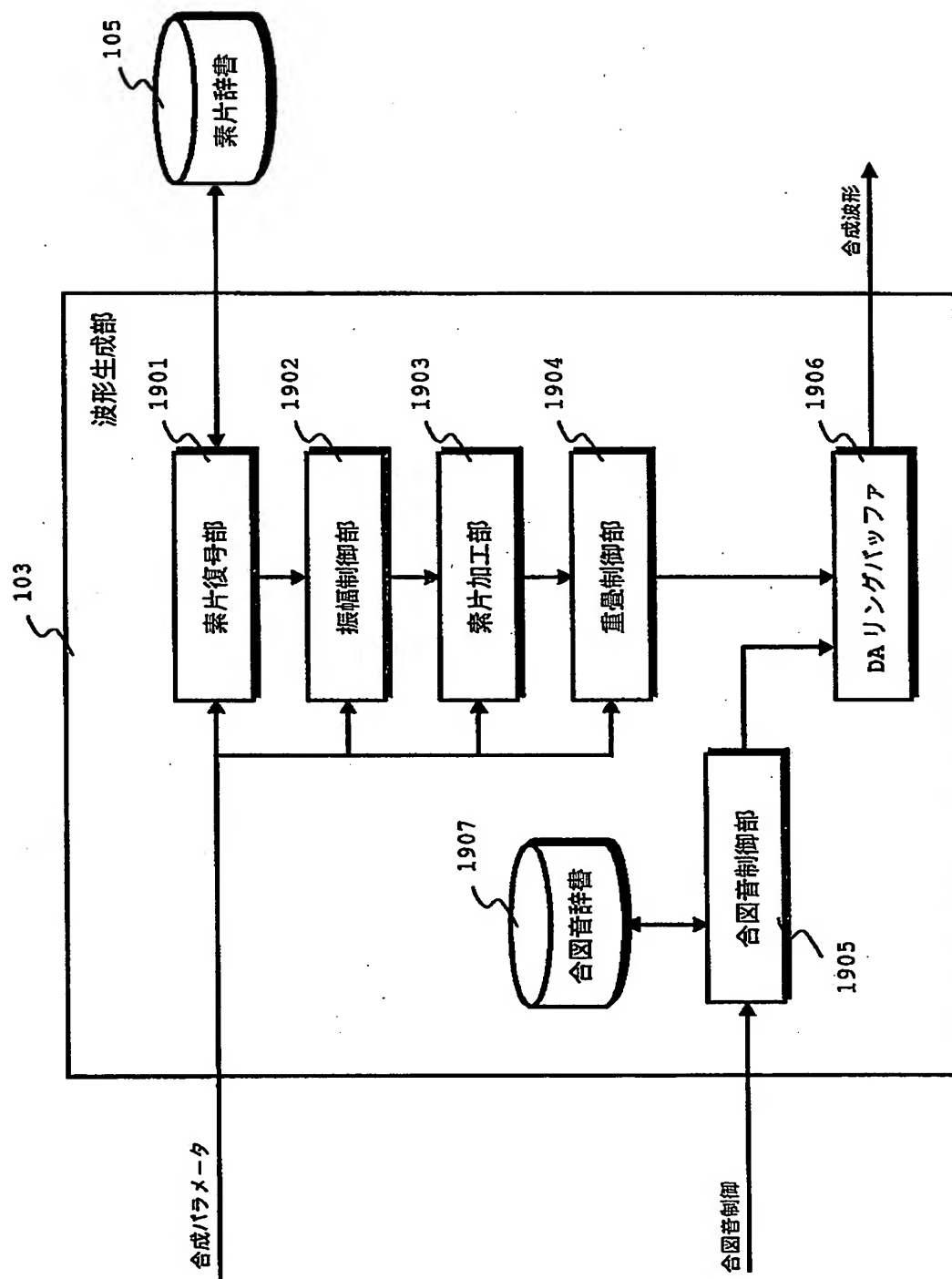
本発明の第 3 の実施の形態におけるパラメータ生成部の機能ブロック

【図 11】



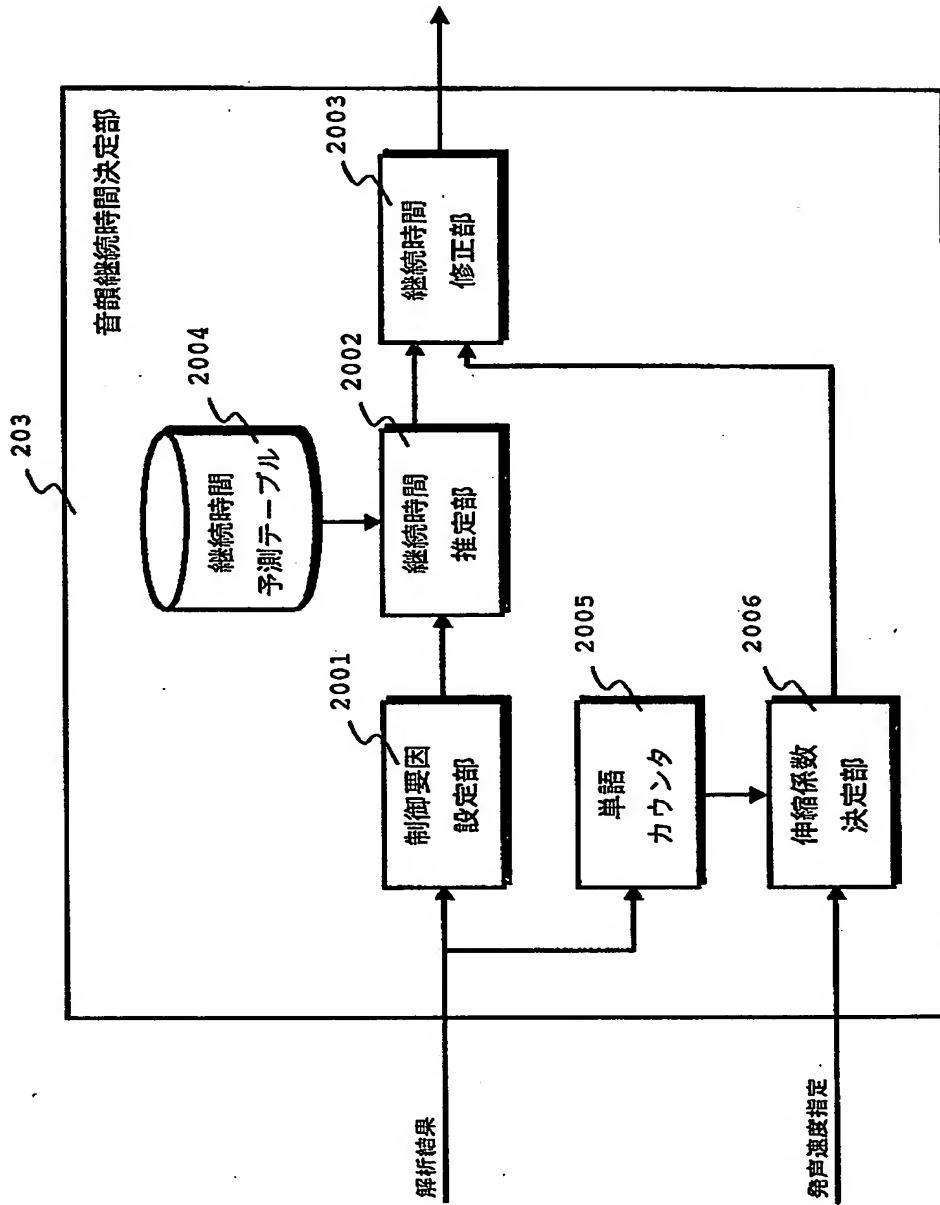
本発明の第 3 の実施の形態における合図音決定部の機能ブロック図

【図 12】



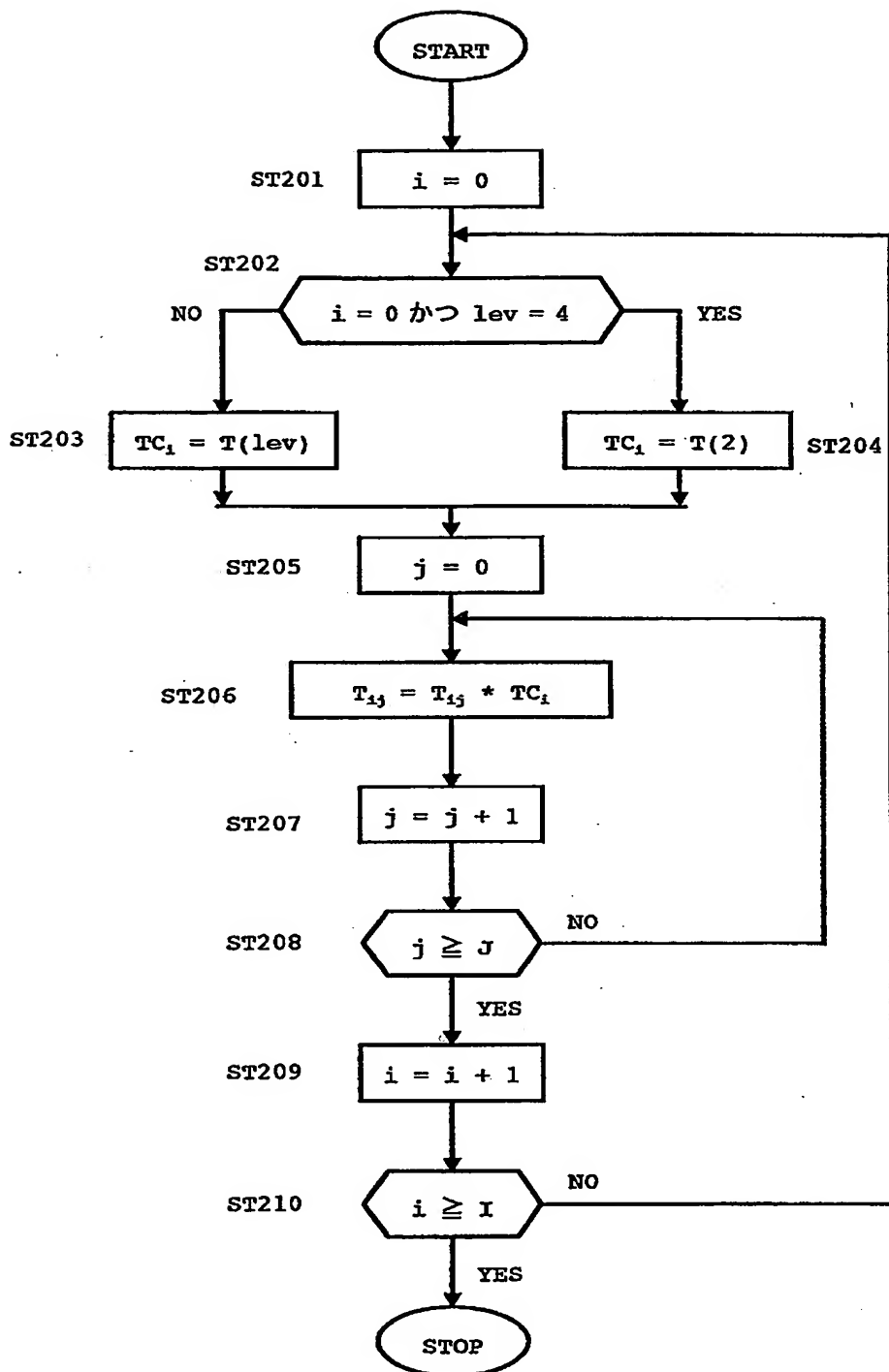
本発明の第 3 の実施の形態における波形生成部の機能ブロック

【図 13】



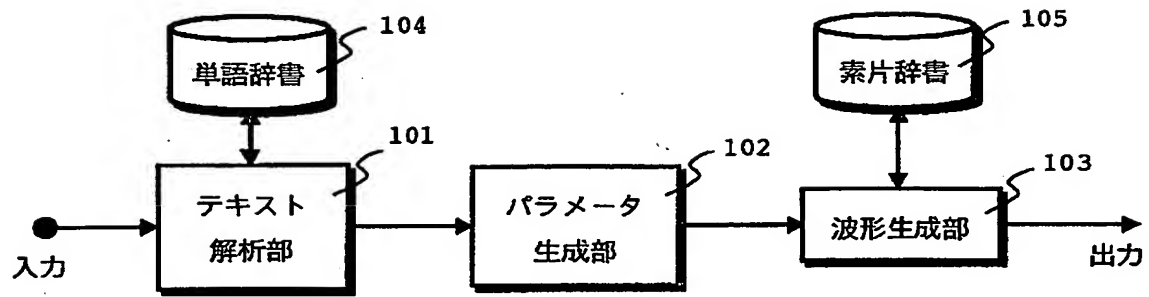
本発明の第 4 の実施の形態における音韻継続時間決定部の機能ブロック

【図 14】



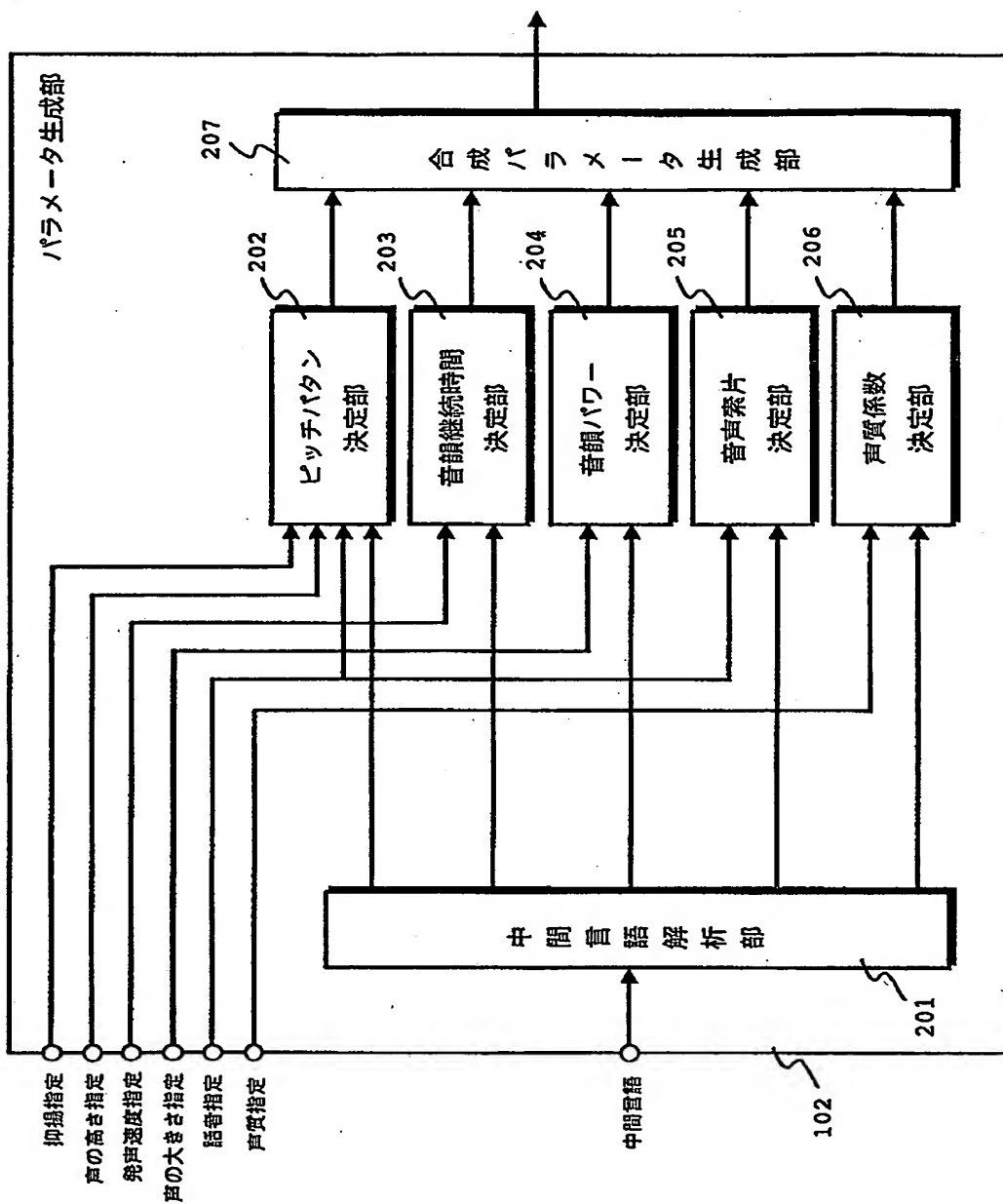
本発明の第 4 の実施の形態における継続時間決定フローチャート

【図 1 5】



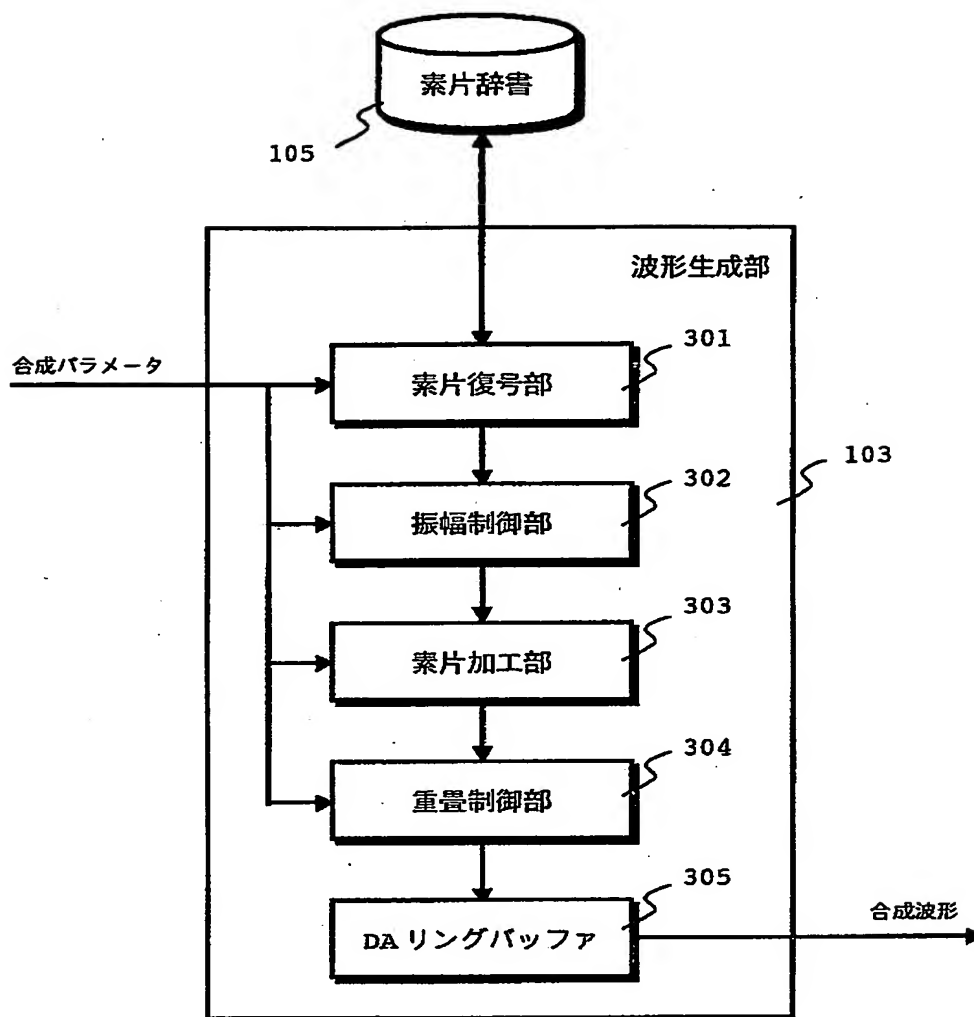
代表的なテキスト音声変換処理の機能ブロック図

【図16】



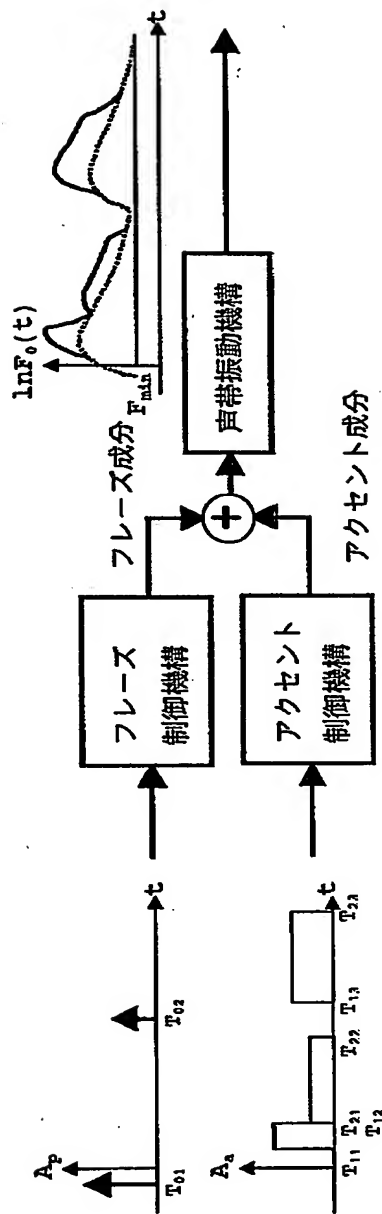
従来技術によるパラメータ生成部の機能ブロック図

【図 1 7】



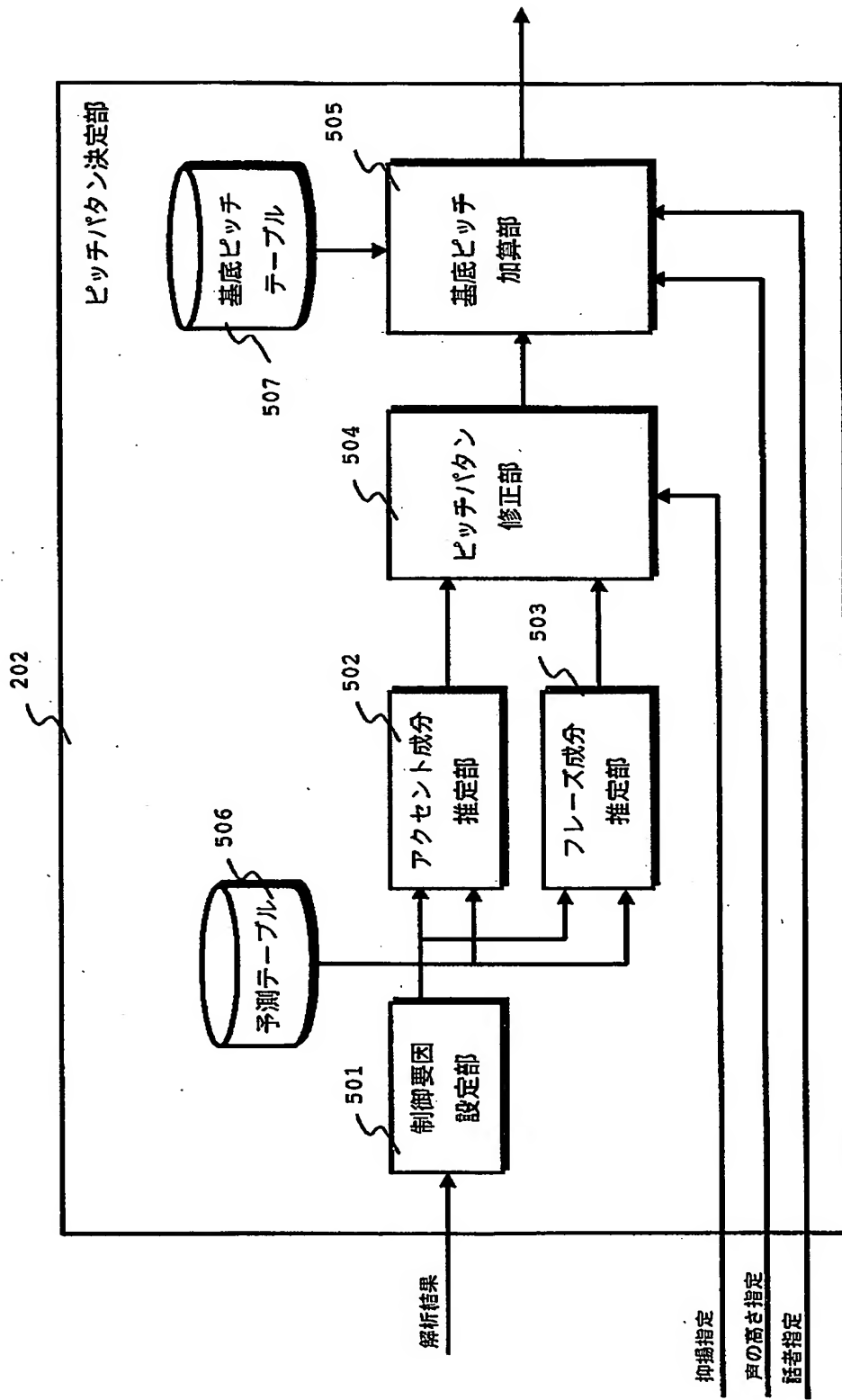
従来技術による波形生成部の機能ブロック図

【図18】



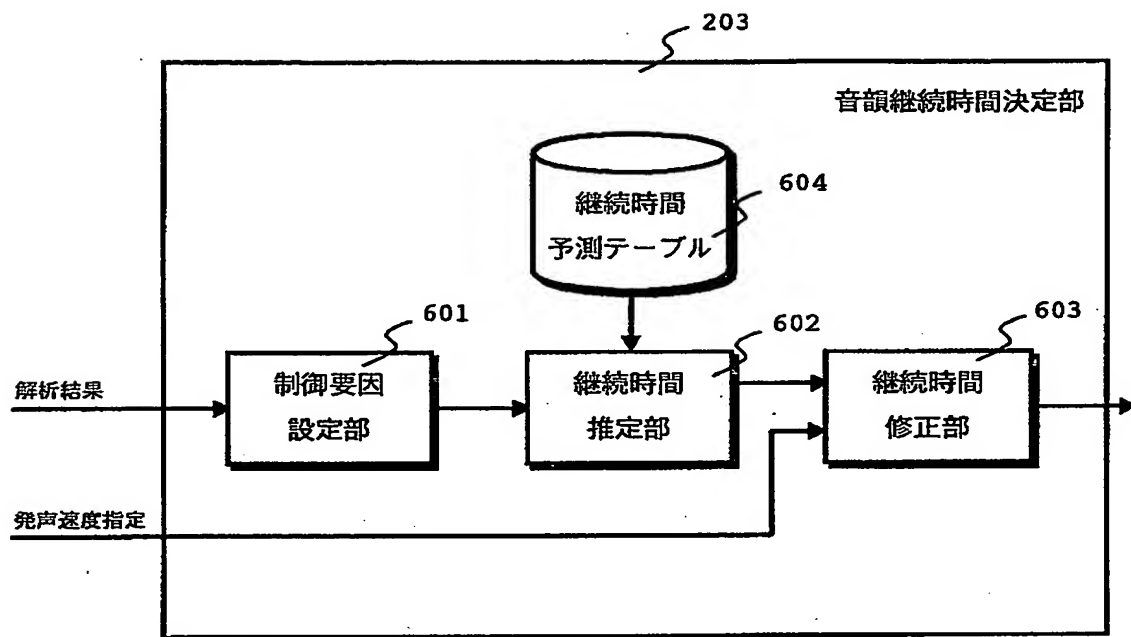
ピッチバタン生成過程モデルの説明図

【図 19】



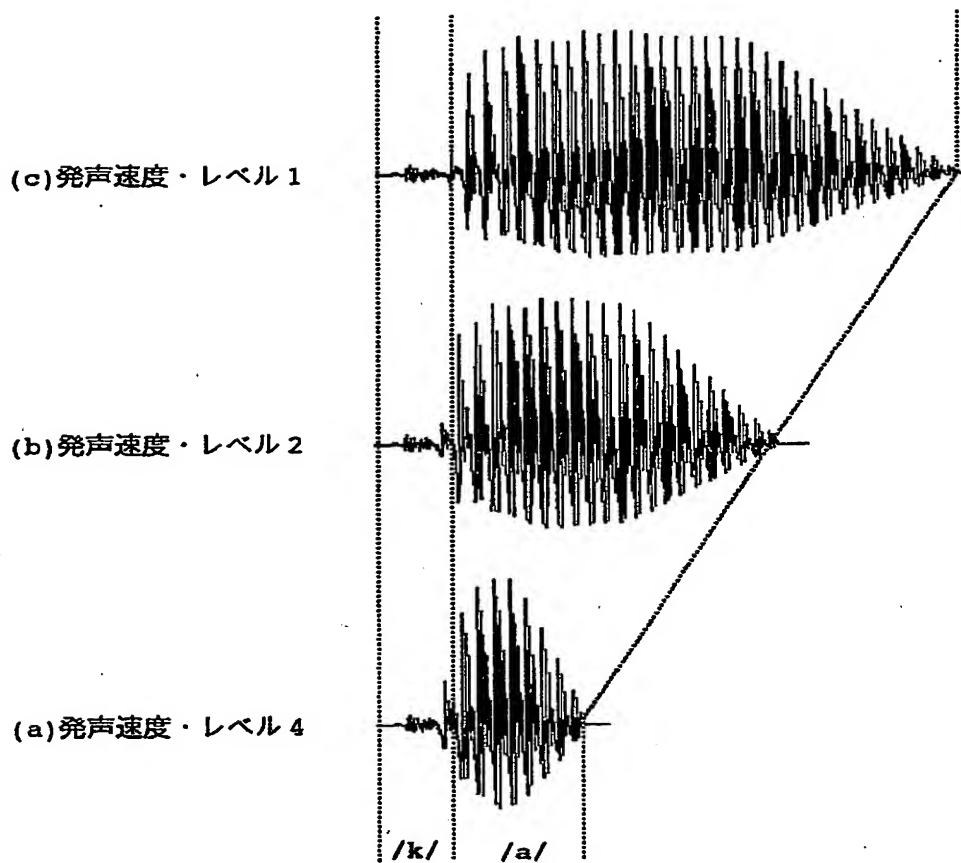
従来技術によるピッチパターン決定部の機能ブロック図

【図 2 0】



従来技術による音韻継続時間決定部の機能ブロック図

【図21】



発声速度の違いによる波形伸縮の説明図

【書類名】 要約書

【要約】

【課題】 発声速度を速くした時に高負荷になって音切れが発生するという問題点を解決したテキスト音声変換装置における高速読み上げ制御方法を提供する。

【解決手段】 テキスト解析部（102）と、音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成部（102）と、素片辞書（105）を参照しながら合成波形を生成する波形生成部（103）とを備え、パラメータ生成部は、音韻継続時間を予め経験的に求めた継続時間規則テーブルと、音韻継続時間を統計的手法を用いて予測した継続時間予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記継続時間規則テーブルを用い、閾値を超えていない時には前記継続時間予測テーブルを用いて音韻継続時間の決定を行う構成とした。

【選択図】 図1

認定・付加情報

特許出願の番号	特願2001-192778
受付番号	50100926763
書類名	特許願
担当官	第一担当上席 0090
作成日	平成13年 6月27日

<認定情報・付加情報>

【提出日】	平成13年 6月26日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [000000295]

1. 変更年月日 1990年 8月22日

[変更理由] 新規登録

住 所 東京都港区虎ノ門1丁目7番12号

氏 名 沖電気工業株式会社